CVPR 2008 Abstracts (Main Conference)

8:30am – 10:30am Poster Session P1A-1: Statistical Methods and Learning (Summit Hall)

1. Face Shape Recovery from a Single Image Using CCA Mapping between Tensor Spaces

Zhen Lei, Qinxun Bai, Ran He, Stan Li

In this paper, we propose a new approach for face shape recovery from a single image. A single near infrared (NIR) image is used as the input, and a mapping from the NIR tensor space to 3D tensor space, learned by using statistical learning, is used for the shape recovery. In the learning phase, the two tensor models are constructed for NIR and 3D images respectively, and a canonical correlation analysis (CCA) based multi-variate mapping from NIR to 3D faces is learned from a given training set of NIR-3D face pairs. In the reconstruction phase, given an NIR face image, the depth map is computed directly using the learned mapping with the help of tensor models. Experimental results are provided to evaluate the accuracy and speed of the method. The work provides a practical solution for reliable and fast shape recovery and modeling of 3D objects.

2. Classifiability-Based Optimal Discriminatory Projection Pursuit

Yu Su, Shiguang Shan, Xilin Chen, Wen Gao

Linear Discriminant Analysis (LDA) might be the most widely used linear feature extraction method in pattern recognition. Based on the analysis on the several limitations of traditional LDA, this paper makes an effort to propose a new computational paradigm named Optimal Discriminatory Projection Pursuit (ODPP), which is totally different from the traditional LDA and its variants. Only two simple steps are involved in the proposed ODPP: one is the construction of candidate projection set; the other is the optimal discriminatory projection pursuit. For the former step, candidate projections are generated as the difference vectors between nearest between-class boundary samples with redundancy well-controlled, while the latter is efficiently achieved by classifiability-based AdaBoost learning from the large candidate projection set. We show that the new “projection pursuit” paradigm not only does not suffer from the limitations of the traditional LDA but also inherits good generalizability from the boundary attribute of candidate projections. Extensive experimental comparisons with LDA and its variants on synthetic and real data sets show that the proposed method consistently has better performances.

3. Blindly Separating Mixtures of Multiple Layers with Spatial Shifts

Kun Gai, Zhenwei Shi, Changshui Zhang

We address the problem of blindly separating mixtures of multiple layer images with unknown spatial shifts and mixing coefficients. Our proposed method can handle the over-determined, determined and under-determined cases where mixtures are more than, as many as and fewer than layers, respectively. The method is fast in over-determined and determined cases, with the same complexity as the fast Fourier transform (FFT), and can separate more layers from fewer mixtures in the under-determined case. It consists of two main steps. First, a novel sparse blind separation algorithm is applied, to estimate the spatial shifts, the mixing coefficients and the edge image of each layer. Second, all layers are reconstructed, by large scale linear programming in the under-determined case, or by least-squares solutions in other cases. The effectiveness of this technology is shown in the experiments on two simulated mixtures of four layers with spatial shifts, real mixture photos containing transparency and reflections, and real mixture images in a dissolve from a video.

4. Structure-Perceptron Learning of a Hierarchical Log-Linear Model

Long Zhu, Yuanhao Chen, Xingyao Ye, Alan Yuille

In this paper, we address the problems of deformable object matching (alignment) and segmentation with cluttered background. We propose a novel hierarchical log-linear model (HLLM) which represents both shape and appearance features at multiple levels of a hierarchy. This model enables us to combine appearance cues at multiple scales directly into the hierarchy and to model shape deformations at short-range, medium range, and long-range. We introduce the *structure-perceptron* algorithm to estimate the parameters of the HLLM in a discriminative way. The learning is able to estimate the appearance and shape parameters simultaneously in a global manner. Moreover, the structure-perceptron learning has a feature selection aspect (similar to AdaBoost) which enables us to specify a class of appearance/shape features and allow the algorithm to select which features to use and weight their importance. This method was applied to the tasks of deformable object localization, segmentation, matching (alignment), and parsing. We demonstrate that the algorithm achieves the state of the art performance by evaluation on public dataset (horse and multi-view face).

5. Unsupervised Learning of Probabilistic Object Models for Object Classification, Segmentation and Recognition

Yuanhao Chen, Long Zhu, Alan Yuille, Hongjiang Zhang

We present a new unsupervised method to learn unified probabilistic object models (POMs) which can be applied to classification, segmentation, and recognition. We formulate this as a structure learning task and our strategy is to learn and combine basic POMs that make use of complementary image cues. Each POM has algorithms for inference and parameter learning, but: (i) the structure of each POM is unknown, and (ii) the inference and parameter learning algorithm for a POM may be impractical without additional information. We address these problems by a novel structure induction procedure which uses *knowledge propagation* to enable POMs to provide information to other POMs and "teach them" (which greatly reduced the amount of supervision required for training). In particular, we learn a POM-IP defined on Interest Points using weak supervision and use this to train a POM-mask, defined on regional features, which yields a combined POM which performs segmentation/localization. This combined model can be used to train POM-edgelets, defined on edgelets, which gives a full POM with improved performance on classification. We give detailed experimental analysis on large datasets which show that the full POM is invariant to scale and rotation of the object (for learning and inference) and performs inference rapidly. In addition, we show that we can apply POMs to learn objects classes (i.e., when there are several objects and the identity of the object in each image is unknown). We emphasize that these models can match between different objects from the same category and hence enable object recognition.

6. Classification via Semi-Riemannian Spaces

Deli Zhao, Zhouchen Lin, Xiaoou Tang

In this paper, we develop a geometric framework for linear or nonlinear discriminant subspace learning and classification. In our framework, the structures of classes are conceptualized as a semi-Riemannian manifold which is considered as a submanifold embedded in an ambient semi-Riemannian space. The class structures of original samples can be characterized and deformed by local metrics of the semi-Riemannian space. Semi-Riemannian metrics are uniquely determined by the smoothing of discrete functions and the nullity of the semi-Riemannian space. Based on the geometrization of class structures, optimizing class structures in the feature space is equivalent to maximizing the quadratic quantities of metric tensors in the semi-Riemannian space. Thus supervised discriminant subspace learning reduces to unsupervised semi-Riemannian manifold learning. Based on the proposed framework, a novel algorithm, dubbed as Semi-Riemannian Discriminant Analysis (SRDA), is presented for subspace-based classification. The performance of SRDA is tested on face recognition (singular case) and handwritten capital letter classification (nonsingular case) against existing algorithms. The experimental results show that SRDA works well on recognition and classification, implying that semi-Riemannian geometry is a promising new tool for pattern recognition and machine learning.

7. Mining Compositional Features for Boosting

Junsong Yuan, Jiebo Luo, Ying Wu

The selection of weak classifiers is critical to the success of boosting techniques. Poor weak classifiers do not perform better than random guess, thus cannot help decrease the training error during the boosting process. Therefore, when constructing the weak classifier pool, we prefer the quality rather than the quantity of the weak classifiers. In this paper, we present a data mining-driven approach to discovering compositional features from a given and possibly small feature pool. Compared with individual features (e.g., weak decision stumps) which are of limited discriminative ability, the mined compositional features have guaranteed power in terms of the descriptive and discriminative abilities, as well as bounded training error. To cope with the combinatorial cost of discovering compositional features, we apply data mining methods (frequent itemset mining) to efficiently find qualified compositional features of any possible order. These weak classifiers are further combined through a multi-class AdaBoost method for final multi-class classification. Experiments on a challenging 10-class event recognition problem show that boosting compositional features can lead to faster decrease of training error and significantly higher accuracy compared to conventional boosting decision stumps.

8. Context-aware Clustering

Junsong Yuan, Ying Wu

Most existing methods of semi-supervised clustering introduce supervision from outside, e.g., manually label some data samples or introduce constrains into clustering results. This paper studies an interesting problem: can the supervision come from inside, i.e., the unsupervised training data themselves? If the data samples are not independent, we can capture the contextual information reflecting the dependency among the data samples, and use it as supervision to improve the clustering. This is called context-aware clustering. The investigation is substantialized on two scenarios of (1) clustering primitive visual features (e.g., SIFT features) with the help of spatial contexts, and (2) clustering `0'-`9' hand written digits with help of contextual patterns among different types of features. Our context-aware clustering can be well formulated in a closed-form, where the contextual information serves as a regularization term to balance the data fidelity in original feature space and the influences of contextual patterns. A nested-EM algorithm is proposed to obtain an efficient solution, which proves to converge. By exploring the dependent structure of the data samples, this method is completely unsupervised, as no outside supervision is introduced.

9. Locally Adaptive Learning for Translation-Variant MRF Image Priors

Masayuki Tanaka, Masatoshi Okutomi

Markov random field (MRF) models are a powerful tool in machine vision applications. However, learning the model parameters is still a challenging problem and a burdensome task. The main contribution of this paper is to propose a locally adaptive learning framework. The proposed learning framework is simple and effective learning framework for translation-variant MRF models. The key idea is to use neighboring patches as a locally adaptive training set. We use multivariate Gaussian MRF models for local image prior models. Although the Gaussian MRF models are too simple for whole natural image priors, the locally adaptive framework enables to express the prior distributions of the every observed image. These locally adaptive learning framework and the multivariate Gaussian translation-variant MRF models simplify the learning procedures. This paper also includes other two contributions; a novel iteration framework by updating the prior information, and a simple and intuitive derivation of the well-known bilateral filter. Experimental results of denoising applications demonstrate that the denoising based on the proposed locally adaptive learning framework outperforms existing high-performance denoising algorithms.

10. Semi-Supervised SVM Batch Mode Active Learning for Image Retrieval

Steven C.H. Hoi, Rong Jin, Jianke Zhu, Micahel R. Lyu

Active learning has been shown as a key technique for improving content-based image retrieval (CBIR) performance. Among various methods, support vector machine (SVM) active learning is popular for its application to relevance feedback in CBIR. However, the regular SVM active learning has two main drawbacks when used for relevance feedback. First, SVM often suffers from learning with a small number of labeled examples, which is the case in relevance feedback. Second, SVM active learning usually does not take into account the redundancy among examples, and therefore could select multiple examples in relevance feedback that are similar (or even identical) to each other. In this paper, we propose a novel scheme that exploits both semi-supervised kernel learning and batch mode active learning for relevance feedback in CBIR. In particular, a kernel function is first learned from a mixture of labeled and unlabeled examples. The kernel will then be used to effectively identify the informative and diverse examples for active learning via a min-max framework. An empirical study with relevance feedback of CBIR showed that the proposed scheme is significantly more effective than other state-of-the-art approaches.

11. Semi-Supervised Distance Metric Learning for Collaborative Image Retrieval

Steven C. H. Hoi, Wei Liu, Shih-Fu Chang

Typical content-based image retrieval (CBIR) solutions with regular Euclidean metric usually cannot achieve satisfactory performance due to the semantic gap challenge. Hence, relevance feedback has been adopted as a promising approach to improve the search performance. In this paper, we propose a novel idea of learning with historical relevance feedback log data, and adopt a new paradigm called "Collaborative Image Retrieval" (CIR). To effectively explore the log data, we propose a novel semi-supervised distance metric learning technique, called "Laplacian Regularized Metric Learning" (LRML), for learning robust distance metrics for CIR. Different from previous methods, the proposed LRML method integrates both log data and unlabeled data information through an effective graph regularization framework. We show that reliable metrics can be learned from real log data even they may be noisy and limited at the beginning stage of a CIR system. We conducted extensive evaluation to compare the proposed method with a large number of competing methods, including 2 standard metrics, 3 unsupervised metrics, and 4 supervised metrics with side information.

12. Multiple-Instance Ranking: Learning to Rank Images for Image Retrieval

Yang Hu, Mingjing Li, Nenghai Yu

We study the problem of learning to rank images for image retrieval. For a noisy set of images indexed or tagged by the same keyword, we learn a ranking model from some training examples and then use the learned model to rank new images. Unlike previous work on image retrieval, which usually coarsely divide the images into relevant and irrelevant images and learn a binary classifier, we learn the ranking model from image pairs with preference relations. In addition to the relevance of images, we are further interested in what portion of the image is of interest to the user. Therefore, we consider images represented by sets of regions and propose multiple-instance rank learning based on the max margin framework. Three different schemes are designed to encode the multiple-instance assumption. We evaluate the performance of the multiple-instance ranking algorithms on real-word images collected from Flickr - a popular photo sharing service. The experimental results show that the proposed algorithms are capable of learning effective ranking models for image retrieval.

13. Correlational Spectral Clustering

Matthew B. Blaschko, Christoph H. Lampert

We present a new method for spectral clustering with paired data based on kernel canonical correlation analysis, called *correlational spectral clustering*. Paired data are common in real world data sources, such as images with text captions. Traditional spectral clustering algorithms either assume that data can be represented by a single similarity measure, or by co-occurrence matrices that are then used in biclustering. In contrast, the proposed method uses separate similarity measures for each data representation, and allows for projection of previously unseen data that are only observed in one representation (e.g., images but not text). We show that this algorithm generalizes traditional spectral clustering algorithms and show consistent empirical improvement over spectral clustering on a variety of datasets of images with associated text.

14. A Parallel Decomposition Solver for SVM: Distributed Dual Ascend using Fenchel Duality

Tamir Hazan, Amit Man, Amnon Shashua

We introduce a distributed algorithm for solving large scale Support Vector Machines (SVM) problems. The algorithm divides the training set into a number of processing nodes each running independently an SVM sub-problem associated with its subset of training data. The algorithm is a parallel (Jacobi) block-update scheme derived from the convex conjugate (Fenchel Duality) form of the original SVM problem. Each update step consists of a modified SVM solver running in parallel over the sub-problems followed by a simple global update. We derive bounds on the number of updates showing that the number of iterations (independent SVM applications on sub-problems) required to obtain a solution of accuracy ε is $O(\log(1/\varepsilon))$. We demonstrate the efficiency and applicability of our algorithms by running on large scale experiments on standardized datasets while comparing the results to the state-of-the-art SVM solvers.

15. High-arity Interactions, Polyhedral Relaxations, and Cutting Plane Algorithm for Soft Constraint Optimisation (MAP-MRF)

Tomas Werner

LP relaxation approach to soft constraint optimisation (i.e., MAP-MRF) has been mostly considered only for binary problems. We present its generalisation to n-ary problems, including a simple algorithm to optimise the LP bound, n-ary max-sum diffusion. As applications, we show that a hierarchy of gradually tighter polyhedral relaxations of MAP-MRF is obtained by adding zero interactions. We propose a cutting plane algorithm, where cuts correspond to adding zero interactions and the separation problem to finding an unsatisfiable constraint satisfaction subproblem. Next, we show that certain high-arity interactions, e.g., certain global constraints, can be included into the framework in a principled way. Finally, we prove that n-ary max-sum diffusion finds global optimum for n-ary supermodular problems.

16. L_1 Regularized Projection Pursuit for Additive Model Learning

Xiao Zhang, Lin Liang, Xiaoou Tang, Heung-Yeung Shum

In this paper, we present a L_1 regularized projection pursuit algorithm for additive model learning. Two new algorithms are developed for regression and classification respectively: sparse projection pursuit regression and sparse Jensen-Shannon Boosting. The introduced L_1 regularized projection pursuit encourages sparse solutions, thus our new algorithms are robust to overfitting and present better generalization ability especially in settings with many irrelevant input features and noisy data. To make the optimization with L_1 regularization more efficient, we develop an "informative feature first" sequential optimization algorithm. Extensive experiments demonstrate the effectiveness of our proposed approach.

17. Semi-Supervised Discriminant Analysis using Robust Path-Based Similarity

Yu Zhang, Dit-Yan Yeung

Linear Discriminant Analysis (LDA), which works by maximizing the within-class similarity and minimizing the between-class similarity simultaneously, is a popular dimensionality reduction technique in pattern recognition and machine learning. In real-world applications when labeled data are limited, LDA does not work well. Under many situations, however, it is easy to obtain unlabeled data in large quantities. In this paper, we propose a novel dimensionality reduction method, called Semi-Supervised Discriminant Analysis (SSDA), which can utilize both labeled and unlabeled data to perform dimensionality reduction in the semisupervised setting. Our method uses a robust path-based similarity measure to capture the manifold structure of the data and then uses the obtained similarity to maximize the separability between different classes. A kernel extension of the proposed method for nonlinear dimensionality reduction in the semi-supervised setting is also presented. Experiments on face recognition demonstrate the effectiveness of the proposed method.

18. Semantic-based Indexing of Fetal Anatomies from 3-D Ultrasound Data Using Global/Semi-local Context and Sequential Sampling

Gustavo Carneiro, Fernando Amat, Bogdan Georgescu, Sara Good, Dorin Comaniciu

The use of 3-D ultrasound data has several advantages over 2-D ultrasound for fetal biometric measurements, such as considerable decrease in the examination time, possibility of post-exam data processing by experts and the ability to produce 2-D views of the fetal anatomies in orientations that cannot be seen in common 2-D ultrasound exams. However, the search for standardized planes and the precise localization of fetal anatomies in ultrasound volumes are hard and time consuming processes even for expert physicians and sonographers. The relative low resolution in ultrasound volumes, small size of fetus anatomies and inter-volume position, orientation and size variability make this localization problem even more challenging. In order to make the plane search and fetal anatomy localization problems completely automatic, we introduce a novel principled probabilistic model that combines discriminative and generative classifiers with contextual information and sequential sampling. We implement a system based on this model, where the user queries consist of semantic keywords that represent anatomical structures of interest. After queried, the system automatically displays standardized planes and produces biometric measurements of the fetal anatomies. Experimental results on a held-out test set show that the automatic measurements are within the inter-user variability of expert users. It resolves for position, orientation and size of three different anatomies in less than 10 seconds in a dual-core computer running at 1.7 GHz.

19. A Hierarchical and Contextual Model for Aerial Image Understanding

Jacob Porway, Kristy Wang, Benjamin Yao, Song Chun Zhu

In this paper we present a novel method for parsing aerial images with a hierarchical and contextual model learned in a statistical framework. We learn hierarchies at the scene and object levels to handle the difficult task of representing scene elements at different scales and add contextual constraints to resolve ambiguities in the scene interpretation. This allows the model to rule out inconsistent detections, like cars on trees, and to verify low probability detections based on their local context, such as small cars in parking lots. We also present a two-step algorithm for parsing aerial images that first detects object-level elements like trees and parking lots using color histograms and bag-of-words models, and objects like roofs and roads using compositional boosting, a powerful method for finding image structures. We then activate the top-down scene model to prune false positives from the first stage. We learn this scene model in a minimax entropy framework and show unique samples from our prior model, which capture the layout of scene objects. We present experiments showing that hierarchical and contextual information greatly reduces the number of false positives in our results.

20. Local Probabilistic Regression for Activity-independent Human Pose Inference

Raquel Urtasun, Trevor Darrell

Discriminative approaches to human pose inference involve mapping visual observations to articulated body configurations. Current probabilistic approaches to learn this mapping have been limited in their ability to handle domains with a large number of activities that require very large training sets. We propose an online probabilistic regression scheme for efficient inference of complex, high-dimensional, and multimodal mappings. Our technique is based on a local mixture of Gaussian Processes, where locality is defined based on both appearance and pose, and where the mapping hyper-parameters can vary across local neighborhoods to better adapt to specific regions in the pose space. The mixture components are defined online in very small neighborhoods, so learning and inference is extremely efficient. When the mapping is one-to-one, we derive a bound on the approximation error of local regression (vs. global regression) for monotonically decreasing covariance functions. Our method can determine when training examples are redundant given the rest of the database, and use this criteria for pruning. We report results on synthetic (Poser) and real (Humaneva) pose databases, obtaining fast and accurate pose estimates using training set sizes up to 10^5.

21. Re-weighting Linear Discrimination Analysis under Ranking Loss

Yong Ma, Yoshihisa Ijiri, Shihong Lao, Masato Kawade

Linear Discrimination Analysis (LDA) is one of the most popular feature extraction and classifier design techniques. It maximizes the Fisher-ratio between the between-class scatter matrix and the within-class scatter matrix under a linear transformation, and the transformation is composed of the generalized eigenvectors of them. However, Fisher criterion itself cannot decide the optimum norm of transformation vectors for classification. In this paper, we show that actually the norm of the transformation vectors has strong influence on classification performance, and we propose a novel method to estimate the optimum norm of LDA under the ranking loss, re-weighting LDA. On artificial data and real databases, the experiments demonstrate the proposed method can effectively improve the performance of LDA classifiers. And the algorithm can also be applied to other LDA variants such as Non-parametric Discriminant Analysis (NDA) to improve their performance further.

22. Latent Topic Random Fields: Learning using a taxonomy of labels

Xuming He, Richard S. Zemel

An important problem in image labeling concerns learning with images labeled at varying levels of specificity. We propose an approach that can incorporate images with labels drawn from a semantic hierarchy, and can also readily cope with missing labels, and roughly-specified object boundaries. We introduce a new form of latent topic model, learning a novel context representation in the joint label-and-image space by capturing co-occurring patterns within and between image features and object labels. Given a topic, the model generates the input data, as well as a topic-dependent probabilistic classifier to predict labels for image regions. We present results on two real-world datasets, demonstrating significant improvements gained by including the coarsely labeled images.

23. Minimal Local Reconstruction Error Measure Based Discriminant Feature Extraction and Classification

Jian Yang, Zhen Lou, Zhong Jin, Jingyu Yang

This paper introduces the minimal local reconstruction error (MLRE) as a similarity measure and presents a MLRE-based classifier. From the geometric meaning of the minimal local reconstruction error, we derive that the MLRE-based classifier is a generalization of the conventional nearest neighbor classifier and the nearest neighbor line and plane classifiers. We further apply the MLRE measure to characterize the within-class and between-class local scatters and then develop a MLRE measure based discriminant feature extraction method. The proposed MLRE-based feature extraction method is in line with the MLRE-based classification method in spirit, thus the two methods can be seamlessly combined in applications. The experimental results on the CENPARMI handwritten numeral database and the FERET face image database show the effectiveness of the proposed MLRE-based feature extraction and classification method.

24. Consistent Image Analogies using Semi-supervised Learning

Li Cheng, S. V. N. Vishwanathan, Xinhua Zhang

In this paper we study the following problem: given two source images A and A', and a target image B, can we learn to synthesize a new image B' which relates to B in the same way that A' relates to A? We propose an algorithm which a) uses a semi-supervised component to exploit the fact that the target image B is available *a priori*, b) uses inference on a Markov Random Field (MRF) to ensure global consistency, and c) uses image quilting to ensure local consistency. Our algorithm can also deal with the case when A is only partially labeled, that is, only small parts of A' are available for training. Empirical evaluation shows that our algorithm consistently produces visually pleasing results, outperforming the state of the art.

25. Learning A Geometry Integrated Image Appearance Manifold From A Small Training Set

Yilei Xu, Amit Roy-Chowdhury

While low-dimensional image representations have been very popular in computer vision, they suffer from two limitations: (i) they require collecting a large and varied training set to learn a low-dimensional set of basis functions, and (ii) they do not retain information about the 3D geometry of the object being imaged. In this paper, we show that it is possible to estimate low-dimensional manifolds that describe object appearance while retaining the geometrical information about the 3D structure of the object. By using a combination of analytically derived geometrical models and statistical learning methods, this can be achieved using a much smaller training set than most of the existing approaches. Specifically, we derive a quadrilinear manifold of object appearance that can represent the effects of illumination, pose, identity and deformation, and the basis functions of the tangent space to this manifold depend on the 3D surface normals of the objects. We show experimental results on constructing this manifold and how to efficiently track on it using an inverse compositional algorithm.

26. Learning Subcategory Relevances to Category Recognition

Sinisa Todorovic, Narendra Ahuja

A real-world object category can be viewed as a characteristic configuration of its parts, that are themselves simpler, smaller (sub)categories. Recognition of a category can therefore be made easier by detecting its constituent subcategories and combing these detection results. Given a set of training images, each labeled by an object category contained in it, we present an approach to learning: (1) Taxonomy defined by recursive sharing of subcategories by multiple image categories; (2) Subcategory relevance as the degree of evidence a subcategory offers for the presence of its parent; (3) Likelihood that the image contains a subcategory; and (4) Prior that a subcategory occurs. The images are represented as points in a feature space spanned by confidences in the occurrences of the subcategories. The subcategory relevances are estimated as weights, necessary to rescale the corresponding axes of the feature space so that the images with the same label are closer to each other than to those with different labels. When a new image is encountered, the learned taxonomy, relevances, likelihoods, and priors are used by a linear classifier to categorize the image. On the challenging Caltech-256 dataset, the proposed approach significantly outperforms the best categorizations reported. This result is significant in that it not only demonstrates the advantages of exploiting subcategory taxonomy for recognition, but also suggests that a feature space spanned by part properties, instead of direct object properties, allows for linear separation of image classes.

27. Structure Learning in Random Fields for Heart Motion Abnormality Detection

Mark Schmidt, Glenn Fung, Murphy Kevin, Romer Rosales

Coronary Heart Disease can be diagnosed by assessing the regional motion of the heart walls in ultrasound images of the left ventricle. Even for experts, ultrasound images are difficult to interpret leading to high intra-observer variability. Previous work indicates that in order to approach this problem, the interactions between the different heart regions and their overall influence on the clinical condition of the heart need to be considered. To do this, we propose a method for jointly learning the structure and parameters of conditional random fields, formulating these tasks as a convex optimization problem. We consider block-L1 regularization for each set of features associated with an edge, and formalize an efficient projection method to find the globally optimal penalized maximum likelihood solution. We perform extensive numerical experiments comparing the presented method with related methods that approach the structure learning problem differently. We verify the robustness of our method on echocardiograms collected in routine clinical practice at one hospital.

28. Learning Bayesian Networks with Qualitative Constraints

Qiang Ji, Yan Tong

Graphical models such as Bayesian Networks (BNs) are being increasingly applied to various computer vision problems. One bottleneck in using BN is that learning the BN model parameters often requires a large amount of reliable and representative training data, which proves to be difficult to acquire for many computer vision tasks. On the other hand, there is often available qualitative prior knowledge about the model. Such knowledge comes either from domain experts based on their experience or from various physical or geometric constraints that govern the objects we try to model. Unlike the quantitative prior, the qualitative prior is often ignored due to the difficulty of incorporating them into the model learning process.

In this paper, we introduce a closed-form solution to systematically combine the limited training data with some generic qualitative knowledge for BN mode parameter learning. To validate our method, we compare it with the Maximum Likelihood (ML) estimation method under sparse data and with the Expectation Maximization (EM) algorithm under incomplete data respectively. To further demonstrate its applications for computer vision, we apply it to learn a BN model for facial Action Unit (AU) recognition from real image data. The experimental results show that with simple and generic qualitative constraints and using only a small amount of training data, our method can robustly and accurately estimate the BN model parameters.

29. Semi-Supervised Learning of Multi-Factor Models for Face De-Identification

Ralph Gross, Latanya Sweeney, Fernando De la Torre, Simon Baker

With the emergence of new applications centered around the sharing of image data, questions concerning the protection of the privacy of people visible in the scene arise. Recently, formal methods for the de-identification of images have been proposed which would benefit from multi-factor coding to separate identity and non-identity related factors. However, existing multi-factor models require complete labels during training which are often not available in practice. In this paper we propose a new multi-factor framework which unifies linear, bilinear, and quadratic models. We describe a new fitting algorithm which jointly estimates all model parameters and show that it outperforms the standard alternating algorithm. We furthermore describe how to avoid over-fitting the model and how to train the model in a semi-supervised manner. In experiments on a large expression-variant face database we show that data coded using our multi-factor model leads to improved data utility while providing the same privacy protection.

30. Incremental Learning of Nonparametric Bayesian Mixture Models

Ryan Gomes, Max Welling, Pietro Perona

Clustering is a fundamental task in many vision applications. To date, most clustering algorithms work in a batch setting and training examples must be gathered in a large group before learning can begin. Here we explore incremental clustering, in which data can arrive continuously. We present a novel incremental model-based clustering algorithm based on nonparametric Bayesian methods, which we call Memory Bounded Variational Dirichlet Process (MB-VDP). The number of clusters are determined flexibly by the data and the approach can be used to automatically discover object categories. The computational requirements required to produce model updates are bounded and do not grow with the amount of data processed. The technique is well suited to very large datasets, and we show that our approach outperforms existing online alternatives for learning nonparametric Bayesian mixture models.

31. Beyond Pairwise Belief Propagation : Labeling by Approximating Kikuchi Free Energies

Ifeoma Nwogu, Jason Corso

Belief Propagation (BP) can be very useful and efficient for performing approximate inference on graphs. But when the graph is very highly connected with strong conflicting interactions, BP tends to fail to converge. Generalized Belief Propagation (GBP) provides more accurate solutions on such graphs, by approximating Kikuchi free energies, but the clusters required for the Kikuchi approximations are hard to generate. We propose a new algorithmic way of generating such clusters from a graph without exponentially increasing the size of the graph during triangulation. In order to perform the statistical region labeling, we introduce the use of super-pixels for the nodes of the graph, as it is a more natural representation of an image than the pixel grid. This results in a smaller but much more highly interconnected graph where BP consistently fails. We demonstrate how our version of the GBP algorithm outperforms BP on synthetic and natural images and, in both cases, GBP converges after only a few iterations.

32. Scene Classification with Low-dimensional Semantic Spaces and Weak Supervision

Nikhil Rasiwasia, Nuno Vasconcelos

A novel approach to scene categorization is proposed. Similar to previous works, we introduce an intermediate space, based on a low dimensional semantic “theme” image representation. However, instead of learning the themes in an unsupervised manner, they are learned with weak supervision, from casual image annotations. Each theme induces a probability density on the space of low-level features, and images are represented as vectors of posterior theme probabilities. This enables an image to be associated with multiple themes, even when there are no multiple associations in the training labels. An implementation is presented and compared to various existing algorithms, on benchmark datasets. It is shown that the proposed low dimensional representation correlates well with human scene understanding, and is able to learn theme co-occurrences without explicit training. It is also shown to outperform unsupervised latent-space methods, with much smaller training complexity, and to achieve performance close to the state of the art methods, which rely on much higher-dimensional image representations. Finally a study of the effect of dimensionality on the classification performance is presented, indicating that the dimensionality of theme space grows sub-linearly with the number of scene categories.

33. Joint Conditional Random Field of Multiple Views with Online Learning for Image-based Rendering

Wenfeng Li, Baoxin Li

There are many applications, such as image-based rendering, where multiple views of a scene are considered simultaneously for improved analysis through employing strong correlation among the set of pixels corresponding to the same physical scene point. While being a useful tool for modeling pixel interactions, Markov Random Field (MRF) models encounter challenges in such cases since they assume strong independence of the observed data for tractability, rendering it difficult to take advantage of having multiple correlated views. In this paper we propose joint Conditional Random Field (CRF) for multiple views in the context of virtual view synthesis in image-based rendering. The model is enabled by the adoption of steerable spatial filters for capturing not only the pixel dependence in a single image but also their correlations among multiple views. Furthermore, a novel on-line learning scheme is proposed for the CRF model, which learns the CRF parameters from the same input data for synthesizing virtual views. This effectively makes the model adaptive to the input and thus optimal results can be expected. Experiments are designed to validate the proposed approach and its effectiveness.

34. Bayesian Tactile Face

Zheshen Wang, Xinyu Xu, Baoxin Li

Computer users with visual impairment cannot access the rich graphical contents in print or digital media unless relying on visual-to-tactile conversion, which is done primarily by human specialists. Automated approaches to this conversion are an emerging research field, in which currently only simple graphics such as diagrams are handled. This paper proposes a systematic method for automatically converting a human portrait image into its tactile form. We model the face based on deformable Active Shape Model (ASM), which is enriched by local appearance models in terms of gradient profiles along the shape. The generic face model including the appearance components is learnt from a set of training face images. Given a new portrait image, the prior model is updated through Bayesian inference. To facilitate the incorporation of a pose-dependent appearance model, we propose a statistical sampling scheme for the inference task. Furthermore, to compensate for the simplicity of the face model, edge segments of a given image are used to enrich the basic face model in generating the final tactile printout. Experiments are designed to evaluate the performance of the proposed method.

35. Unified Principal Component Analysis with Generalized Covariance Matrix for Face Recognition

Shiguang Shan, Bo Cao, Yu Su, Laiyun Qing, Xilin Chen, Wen Gao

Recently, 2DPCA and its variants have attracted much attention in the face recognition area. In this paper, some efforts are made to discover the underlying fundaments of these methods, and a novel framework called Unified Principal Component Analysis (UPCA) is proposed. First, we introduce a novel concept, named Generalized Covariance Matrix (GCM), which is naturally derived from the traditional Covariance Matrix (CM). Each element of GCM is a generalized covariance of two random vectors rather than two scalar variables in CM. Based on GCM, the UPCA framework is proposed, from which the traditional PCA and its 2D counterparts can be deduced as special cases. Furthermore, under the UPCA framework, we not only revisit the existing 2D PCA methods and their limitations, but also propose two new methods: the grid-sampling method (GridPCA) and the intra-group correlation reduction method. Extensive experimental results on the FERET face database support the theoretical analysis and validate the feasibility of the proposed methods.

36. Discovering Class Specific Compositive Features through Discriminative Sampling with Swendsen-Wang Cut

Feng Han, Ying Shan, Harpreet Sawhney, Rakesh Kumar

This paper proposes a novel approach to discover a set of class specific "composite features" as the feature pool for the detection and classification of complex objects using AdaBoost. Each composite feature is constructed from the combination of multiple individual features. Unlike previous works that design features manually or with certain restrictions, the class specific features are selected from the space of all combinations of a set of individual features. To achieve this, we first establish an analogue between the problem of discriminative feature selection and generative image segmentation, and then draw discriminative samples from the combinatory space with a novel algorithm called *Discriminative Generalized Swendsen-Wang Cut*. These samples form the initial pool of features, where AdaBoost is applied to learn a strong classifier combining the most discriminative composite features. We demonstrate the efficacy of our approach by comparing with existing detection algorithms for finding people in general pose.

37. A Unified Framework for Generalized Linear Discriminant Analysis

Shuiwang Ji, Jieping Ye

Linear Discriminant Analysis (LDA) is one of the well-known methods for supervised dimensionality reduction. Over the years, many LDA-based algorithms have been developed to cope with the curse of dimensionality. In essence, most of these algorithms employ various techniques to deal with the singularity problem, which occurs when the data dimensionality is larger than the sample size. They have been applied successfully in various applications. However, there is a lack of a systematic study of the commonalities and differences of these algorithms, as well as their intrinsic relationships. In this paper, a unified framework for generalized LDA is proposed via a transfer function. The proposed framework elucidates the properties of various algorithms and their relationships. Based on the presented analysis, we propose an efficient model selection algorithm for LDA. We conduct extensive experiments using a collection of high-dimensional data, including text documents, face images, gene expression data, and gene expression pattern images, to evaluate the proposed theories and algorithms.

38. A Deformable Local Image Descriptor

Hong Cheng, Zicheng Liu, Nanning Zheng, Jie Yang

This paper presents a novel local image descriptor that is robust to general image deformations. A limitation with traditional image descriptors is that they use a single support region for each interest point. For general image deformations, the amount of deformation for each location varies and is unpredictable such that it is difficult to choose the best scale of the support region. To overcome this difficulty, we propose to use multiple support regions of different sizes surrounding an interest point. A feature vector is computed for each support region, and the concatenation of these feature vectors forms the descriptor for this interest point. Furthermore, we propose a new similarity measure model, Local-to-Global Similarity (LGS) model, for point matching that takes advantage of the multi-size support regions. Each support region acts as a 'weak' classifier and the weights of these classifiers are learned in an unsupervised manner. The proposed approach is evaluated on a number of images with real and synthetic deformations. The experiment results show that our method outperforms existing techniques under different deformations.

39. A Joint Appearance-Spatial Distance for Kernel-Based Image Categorization

Guo-Jun Qi, Xian-Sheng Hua, Yong Rui, Jinhui Tang, Zheng-Jun Zha, Hong-Jiang Zhang

The goal of image categorization is to classify a collection of unlabeled images into a set of predefined classes to support semantic-level image retrieval. The distance measures used in most existing approaches either ignored the spatial structures or used them in a separate step. As a result, these distance measures achieved only limited success. To address these difficulties, in this paper, we propose a new distance measure that integrates joint appearance-spatial image features. Such a distance measure is computed as an upper bound of an information-theoretic discrimination, and can be computed efficiently in a recursive formulation that scales well to image size. In addition, the upper bound approximation can be further tightened via adaption learning from a universal reference model. Extensive experiments on two widely-used data sets show that the proposed approach significantly outperforms the state-of-the-art approaches.

40. A Learning-based Hybrid Tagging and Browsing Approach for Efficient Manual Image Annotation

Rong Yan, Apostol Natsev, Murray Campbell

In this paper we introduce a learning approach to improve the efficiency of manual image annotation. Although important in practice, manual image annotation has rarely been studied in a quantitative way. We propose formal models to characterize the annotation times for two commonly used manual annotation approaches, i.e., tagging and browsing. The formal models make clear the complementary properties of these two approaches, and inspire a learning-based hybrid annotation algorithm. Our experiments show that the proposed algorithm can achieve up to a 50% reduction in annotation time over baseline methods.

41. Spectral Methods for Semi-supervised Manifold Learning

Zhenyue Zhang, Hongyuan Zha, Min Zhang

Given a finite number of data points sampled from a low-dimensional manifold embedded in a high dimensional space together with the parameter vectors for a subset of the data points, we need to determine the parameter vectors for the rest of the data points. This problem is known as semi-supervised manifold learning, and in this paper we propose methods to handle this problem by solving certain eigenvalue-problems. Our proposed methods address two key issues in semi-supervised manifold learning: 1) fitting of the local affine geometric structures, and 2) preserving the global manifold structures embodied in the overlapping neighborhoods around each data points. We augment the alignment matrix of local tangent space alignment (LTSA) with the orthogonal projection based on the known parameter vectors, giving rise to the eigenvalue problem that characterizes the semi-supervised manifold learning problem. We also discuss the roles of different types of neighborhoods and their influence on the learning process. We illustrate the performance of the proposed methods using both synthetic data sets as well as data sets arising from applications in video annotations.

42. Annotating Collections of Geotagged Photos Using Hierarchical Event and Scene Models

Liangliang Cao, Jiebo Luo, Henry Kautz, Thomas Huang

Most image annotation systems consider a single photo at a time and label photos individually. In this work, we focus on collections of personal photos and explore the associated GPS and time information for semantic annotation. First, we employ a constrained clustering method to partition a photo collection into event-based sub-collections, considering that the GPS records may be partly missing (a practical issue). We then use conditional random field (CRF) models to exploit the correlation between photos based on (1) time-location constraints and (2) the relationship between collection-level annotation (i.e., events) and image-level annotation (i.e., scenes). With the introduction of such a multi-level annotation hierarchy, our system addresses the problem of annotating consumer photo collections that requires a more hierarchical description of the customers' activities than do the simpler image annotation tasks. The efficacy of the proposed system is validated using a geotagged customer photo collection database, which consists of over 100 folders and is labeled for 12 events and 12 scenes.

43. Two Dimensional Active Learning for Image Classification

Guo-Jun Qi, Xian-Sheng Hua, Yong Rui, Jinhui Tang, Hong-Jiang Zhang

In this paper, we propose a two-dimensional active learning scheme and show its application in image classification. Traditional active learning methods select samples only along the sample dimension. While this is the right strategy in binary classification, it is sub-optimal for multilabel classification. In multi-label classification, we argue that, for each selected sample, only a part of more effective labels are necessary to be annotated while others can be inferred by exploring the correlations among the labels. The reason is that the contributions of different labels to minimizing the classification error are different due to the inherent label correlations. To this end, we propose to select sample-label pairs, rather than only samples, to minimize a multi-label Bayesian classification error bound. This new active learning strategy not only considers the sample dimension but also the label dimension, and we call it Two-Dimensional Active Learning (2DAL). We also show that the traditional active learning formulation is a special case of 2DAL when there is only one label. Extensive experiments conducted on two real-world applications show that the 2DAL significantly outperforms the best existing approaches which did not take label correlation into account.

44. Joint Multi-Label Multi-Instance Learning for Image Classification

Zheng-Jun Zha, Xian-Sheng Hua, Mei Tao, Jingdong Wang, Guo-Jun Qi, Zengfu Wang

In real world, an image is usually associated with multiple labels which are characterized by different regions in the image. Thus image classification is naturally posed as both a multi-label learning and multi-instance learning problem. Different from existing research which has considered these two problems separately, we propose an integrated multi-label multi-instance learning (MLMIL) approach based on hidden conditional random fields (HCRFs), which simultaneously captures both the connections between semantic labels and regions, and the correlations among the labels in a single formulation. We apply this MLMIL framework to image classification and report superior performance compared to key existing approaches over the MSR Cambridge (MSRC) and Corel data sets.

45. Pattern Discovery in Motion Time Series via Structure-based Spectral Clustering

Xiaozhe Wang, Liang Wang, Anthony Wirth

This paper proposes an approach called `structure-based spectral clustering' to identify clusters in motion time series for sequential pattern discovery. The proposed approach deploys a `statistical feature-based distance computation' for spectral clustering algorithm. Compared to traditional spectral clustering approaches, in which the similarity matrix is constructed from the original data points by applying some similarity functions, the proposed approach builds the matrix based on a finite set of feature vectors. When the proposed approach uses less data points and simpler similarity function to computing the similarity matrix input for spectral clustering, it can improve the computational efficiency in constructing the similarity graph in spectral clustering compared to conventional approach. Promising experimental results with high accuracy on real world data sets demonstrate the capability and effectiveness of the proposed approach for pattern discovery in motion video sequences.

46. Coherent Image Annotation by Learning Semantic Distance

Tao Mei, Yong Wang, Xian-Sheng Hua, Shaogang Gong, Shipeng Li

Conventional approaches to automatic image annotation usually suffer from two problems: (1) They cannot guarantee a good semantic coherence of the annotated words for each image, as they treat each word independently without considering the inherent semantic coherence among the words; (2) They heavily rely on visual similarity for judging semantic similarity. To address the above issues, we propose a novel approach to image annotation which simultaneously learns a semantic distance by capturing the prior annotation knowledge and propagates the annotation of an image as a whole entity. Specifically, a semantic distance function (SDF) is learned for each semantic cluster to measure the semantic similarity based on relative comparison relations of prior annotations. To annotate a new image, the training images in each cluster are ranked according to their SDF values with respect to this image and their corresponding annotations are then propagated to this image as a whole entity to ensure semantic coherence. We evaluate the innovative SDF-based approach on Corel images compared with Support Vector Machine-based approach. The experiments show that SDF-based approach outperforms in terms of semantic coherence, especially when each training image is associated with multiple words.

47. A Quasi-random Sampling Approach to Image Retrieval

Jun Zhou, Antonio Robles-Kelly

In this paper, we present a novel approach to contents-based image retrieval. The method hinges in the use of quasi-random sampling to retrieve those images in a database which are related to a query image provided by the user. Departing from random sampling theory, we make use of the EM algorithm so as to organize the images in the database into compact clusters that can then be used for stratified random sampling. For the purposes of retrieval, we use the similarity between the query and the clustered images to govern the sampling process within clusters. In this way, the sampling can be viewed as a stratified sampling one which is random at the cluster level and takes into account the intra-cluster structure of the dataset. This approach leads to a measure of statistical confidence that relates to the theoretical hard-limit of the retrieval performance. We show results on the Oxford Flowers dataset.

48. Private Content Based Image Retrieval

Jagarlamudi Shashank, Palivela Kowshik, Kannan Srinathan, C. V. Jawahar

For content level access, very often database needs the query as a sample image. However, the image may contain private information and hence the user does not wish to reveal the image to the database. *Private Content Based Image Retrieval* (PCBIR) deals with retrieving similar images from an image database without revealing the content of the query image--not even to the database server. We propose algorithms for PCBIR, when the database is indexed using hierarchical index structure or hash based indexing scheme. Experiments are conducted on real datasets with popular features and state of the art data structures. It is observed that specialty and subjectivity of image retrieval (unlike SQL queries to a relational database) enables in computationally efficient yet private solutions.

49. Rank-based Distance Metric Learning: An Application to Image Retrieval

Jung-Eun Lee, Rong Jin, Anil Jain

We present a novel approach to learn distance metric for information retrieval. Learning distance metric from a number of queries with side information, i.e., relevance judgements, has been studied widely, for example pair-wise constraint-based distance metric learning. However, the capacity of existing algorithms is limited, because they usually assume that the distance between two similar objects is smaller than the distance between two dissimilar objects. This assumption may not hold, especially in the case of information retrieval when the input space is heterogeneous. To address this problem explicitly, we propose rank-based distance metric learning. Our approach overcomes the drawback of existing algorithms by comparing the distances only among the relevant and irrelevant objects for a given query. To avoid over-fitting, a regularizer based on the Burg matrix divergence is also introduced. We apply the proposed framework to tattoo image retrieval in forensics and law enforcement application domain. The goal of the application is to retrieve tattoo images from a gallery database that are visually similar to a tattoo found on a suspect or a victim. The experimental results show encouraging results in comparison to the standard approaches for distance metric learning.

50. Geo-located image analysis using latent representations

Marco Cristani, Alessandro Perina, Umberto Castellani, Vittorio Murino

Image categorization is undoubtedly one of the most challenging open problems faced in Computer Vision, far from being solved by employing pure visual cues. Recently, additional textual “tags” can be associated to images, enriching their semantic interpretation beyond the pure visual aspect, and helping to bridge the so-called semantic gap. One of the latest class of tags consists in geo-location data, containing information about the geographical site where an image has been captured. Such data motivate, if not require, novel strategies to categorize images, and pose new problems to focus on. In this paper, we present a statistical method for geo-located image categorization, in which categories are formed by clustering geographically proximal images with similar visual appearance. The proposed strategy permits also to deal with the geo-recognition problem, i.e., to infer the geographical area depicted by images with no available location information. The method lies in the wide literature on statistical latent representations; in particular, the probabilistic Latent Semantic Analysis (pLSA) paradigm has been extended, introducing a latent aspect which characterizes peculiar visual features of different geographical zones. Experiments on categorization and georecognition have been carried out employing a well-known geographical image repository: results are actually very promising, opening new interesting challenges and applications in this research field.

51. An Efficient Algorithm for Compressed MR Imaging using Total Variation and Wavelets

Shiqian Ma, Wotao Yin, Yin Zhang, Amit Chakraborty

Compressed sensing, an emerging multidisciplinary field involving mathematics, probability, optimization, and signal processing, focuses on reconstructing an unknown signal from a very limited number of samples. Because information such as boundaries of organs is very sparse in most MR images, compressed sensing makes it possible to reconstruct the same MR image from a very limited set of measurements significantly reducing the MRI scan duration. In order to do that however, one has to solve the difficult problem of minimizing non-smooth functions on large data sets. To handle this, we propose an efficient algorithm that jointly minimizes the ℓ_1 norm, total variation, and a least squares measure, one of the most powerful models for MR images. Our algorithm is based upon an iterative operator-splitting framework. The calculations are accelerated by continuation and takes advantage of fast wavelet and Fourier transforms enabling our code to process MR images from actual real life applications. We show that faithful MR images can be reconstructed from a subset that represents a mere 20 percent of the complete set of measurements.

52. Robust Tensor Factorization Using R1-Norm

Heng Huang, Chris Ding

Over the years, many tensor based algorithms, e.g., two dimensional principle component analysis (2DPCA), two dimensional singular value decomposition (2DSVD), high order SVD, have been proposed for the study of high dimensional data in a large variety of computer vision applications. An intrinsic limitation of previous tensor reduction methods is the sensitivity to the presence of outliers, because they minimize the sum of squares errors (L_2 norm). In this paper, we propose a novel robust tensor factorization method using R_1 norm for error accumulation function using robust covariance matrices, allowing the method to be efficiently implemented instead of resorting to quadratic programming software packages as in other L_1 norm approaches. Experimental results on face representation and reconstruction show that our new robust tensor factorization method can effectively handle outliers compared to previous tensor based PCA methods.

53. Hierarchical, Learning-based Automatic Liver Segmentation

Haibin Ling, S. Kevin Zhou, Yefeng Zheng, Bogdan Georgescu, Michael Suehling, Dorin Comaniciu

In this paper we present a hierarchical, learning-based approach for automatic and accurate liver segmentation from 3D CT volumes. We target CT volumes that come from largely diverse sources (e.g., diseased in six different organs) and are generated by different scanning protocols (e.g., contrast and non-contrast, various resolution and position). Three key ingredients are combined to solve the segmentation problem. First, a hierarchical framework is used to efficiently and effectively monitor the accuracy propagation in a coarse-to-fine fashion. Second, two new learning techniques, marginal space learning and steerable features, are applied for robust boundary inference. This enables handling of highly heterogeneous texture pattern. Third, a novel shape space initialization is proposed to improve traditional methods that are limited to similarity transformation. The proposed approach is tested on a challenging dataset containing 174 volumes. Our approach not only produces excellent segmentation accuracy, but also runs about fifty times faster than state-of-the-art solutions.

54. A Statistical Deformation Prior for Non-Rigid Image and Shape Registration

Thomas Albrecht, Marcel Lüthi, Thomas Vetter

Non-rigid registration is central to many problems in computer vision and medical image analysis. We propose a registration algorithm which is regularized by prior knowledge in the form of a statistical deformation model. This model is obtained from previous registrations performed on a set of noise-free training examples given by images, or shapes represented by level set functions. Contrary to similar approaches, our method does not strictly constrain the result to lie in the span of the statistical model but rather uses the model for Tikhonov regularization. Therefore, our method can be used to reduce the influence of noise and artifacts even when the model contains only a few typical examples. This automatically gives rise to a bootstrapping strategy for building statistical models from noisy data sets requiring only a limited number of high quality examples. We demonstrate the effectiveness of the approach on synthetic and medical images.

55. Computing Minimal Deformations: Application to Construction of Statistical Shape Models

Darko Zikic, Michael Sass Hansen, Glocker Ben, Ali Khamene, Rasmus Larsen, Nassir Navab

Nonlinear registration is mostly performed after initialization by a global, linear transformation (in this work, we focus on similarity transformations), computed by a linear registration method. For the further processing of the results, it is mostly assumed that this preregistration step completely removes the respective linear transformation. However, we show that in deformable settings, this is not the case. As a consequence, a significant linear component is still existent in the deformation computed by the nonlinear registration algorithm. For construction of statistical shape models (SSM) from deformations, this is an unwanted property: SSMs should not contain similarity transformations, since these do not capture information about shape. We propose a method which performs an a posteriori extraction of a similarity transformation from a given nonlinear deformation field, and we use the processed fields as input for SSM construction. For computation of minimal displacements, a closed-form solution minimizing the squared Euclidean norm of the displacement field subject to similarity parameters is used. Experiments on real inter-subject data and on a synthetic example show that the theoretically justified removal of the similarity component by the proposed method has a large influence on the shape model and significantly improves the results.

56. Learning Based Coarse-to-fine Image Registration

Jiang Jiayan, Zheng Songfeng, Toga Arthur, Tu Zhuowen

This paper describes a coarse-to-fine learning based image registration algorithm which has particular advantages in dealing with multi-modality images. Many existing image registration algorithms use a few designed terms or mutual information to measure the similarity between image pairs. Instead, we push the learning aspect by selecting and fusing a large number of features for measuring the similarity. Moreover, the similarity measure is carried in a coarse-to-fine strategy: global similarity measure is first performed to roughly locate the component, we then learn/compute similarity on the local image patches to capture the fine level information. When estimating the transformation parameters, we also engage a coarse-to-fine strategy. Off-the-shelf interest point detectors such as SIFT have degraded results on medical images. We further push the learning idea to extract the main structures/landmarks. Our algorithm is illustrated on three applications: (1) registration of mouse brain images of different modalities, (2) registering human brain image of MRI T1 and T2 images, (3) faces of different expressions. We show greatly improved results over the existing algorithms based on either mutual information or geometric structures.

57. Fully Automatic Feature Localization for Medical Images using a Global Vector Concentration Approach

Tatsuo Kozakaya, Tomoyuki Shibata, Tomoyuki Takeguchi, Masahide Nishiura

In this paper, we propose a novel feature localization method based on a global vector concentration approach. Our approach does not rely on the detection of local salient features around feature points. Instead, we exploit global structural information of the object extracted by calculating the concentration of directional vectors from sampling points. Those vectors are combined with local pattern descriptors of a query image and selected from preliminarily trained extended templates by nearest neighbor search. Due to the insensitivity of local changes, our method can handle partially occluded and noisy objects. We apply the proposed method to fully automatic feature localization of the left ventricular in echocardiograms. The results show the effectiveness of our method in comparison with a conventional edge-based method in terms of accuracy and robustness.

58. Cell Motion Analysis Without Explicit Tracking

Richard Souvenir, Jerrod P. Kraftchick, Sangho Lee, Mark G. Clemens, Min C. Shin

Automated cell tracking using in vivo imagery is difficult, in general, due to the noise inherent in the imaging process, occlusions, varied cell appearance over time, motion of other tissue (distractors), and cells traveling in and out of the image plane. For certain types of cells these problems are exacerbated due to erratic motion patterns. In this paper, we introduce the Radial Flow Transform, which provides motion estimates for objects of interest in a scene without explicitly tracking each object. The transform is robust to misdetected objects, temporally-disjoint motion events, and can represent multiple directions of flow at a single location. We provide operations to convert to and from a vector field representation. This allows for intuitive reasoning about the motion patterns in a scene. We demonstrate results on synthetic data and in vivo microscopy video of a mouse liver.

59. The Statistical Modelling of Fingerprint Minutiae Distribution with Implications for Fingerprint Individuality Studies

Jiansheng Chen, Yiu-Sang Moon

The spatial distribution of fingerprint minutiae is a core problem in the fingerprint individuality study, the cornerstone of the fingerprint authentication technology. Previously, the assumption in most research that minutiae distribution is random has been proved to be inaccurate and may lead to significant overestimates of fingerprint uniqueness. In this paper, we propose a stochastic model for describing and simulating fingerprint minutiae patterns. Through coupling a pair potential Markov point process with a thinned process, this model successfully depicts the complex statistical behavior of fingerprint minutiae. Parameters of this model can be determined by nonlinear minimization. Furthermore, experiment results show that the statistical properties of our proposed model dovetails nicely with real minutiae data in terms of the false fingerprint correspondence probability. Such evidences indicate that the proposed model is a more accurate foundation for minutiae based fingerprint individuality studies as well as the artificial fingerprint synthesis when compared to the model of random distribution.

60. Enforcing Stochastic Inverse Consistency in Non-Rigid Image Registration and Matching

Sai-Kit Yeung, Chi-Keung Tang, Pengcheng Shi, Josien P.W. Pluim, Max A. Viergever, Albert C.S. Chung

This paper presents a new method to enforce inverse consistency in nonrigid image registration and matching. Conventional approaches assume diffeomorphic transformation, implicitly or explicitly. However, the inherent smoothness constraint discourages discontinuity consideration. We propose a *post*-processing algorithm that integrates the input forward and backward fields, which are output by existing registration/matching algorithms, to produce more robust results. Given such a pair of input fields, our algorithm alternately refines the fields by tensor belief propagation, and enforces inverse consistency in stochastic sense by generalized total least squares fitting. To show the efficacy of our stochastic inverse consistency approach, we first present results on very noisy fields. We then demonstrate improvement on existing stereo matching where occlusion is naturally handled by localizing violations of inverse consistency. Finally, we propose a novel application on image stitching, where stochastic inverse consistency is employed in structure deformation, in order to seamlessly align overlapping images with severe misalignment in structure and intensity.

61. Exact Inference in Multi-label CRFs with Higher Order Cliques

Srikumar Ramalingam, Pushmeet Kohli, Karteek Alahari, Philip H. S. Torr

This paper addresses the problem of exactly inferring the maximum a posteriori solutions of discrete multi-label MRFs or CRFs with higher order cliques. We present a framework to transform special classes of multi-label higher order functions to submodular second order boolean functions (referred to as $\mathcal{F}_s^2$), which can be minimized exactly using graph cuts and we characterize those classes. The basic idea is to use two or more boolean variables to encode the states of a single multi-label variable. There are many ways in which this can be done and much interesting research lies in finding ways which are optimal or minimal in some sense. We study the space of possible encodings and find the ones that can transform the most general class of functions to $\mathcal{F}_s^2$. Our main contributions are two-fold. First, we extend the subclass of submodular energy functions that can be minimized exactly using graph cuts. Second, we show how higher order potentials can be used to improve single view 3D reconstruction results. We believe that our work on exact minimization of higher order energy functions will lead to similar improvements in solutions of other labelling problems.

62. Reduce, Reuse & Recycle: Efficiently Solving Multi-Label MRFs

Karteek Alahari, Pushmeet Kohli, Philip H. S. Torr

In this paper, we present novel techniques that improve the computational and memory efficiency of algorithms for solving multi-label energy functions arising from discrete MRFs or CRFs. These methods are motivated by the observations that the performance of minimization algorithms depends on: (a) the initialization used for the primal and dual variables; and (b) the number of primal variables involved in the energy function. Our first method (dynamic alpha-expansion) works by `recycling' results from previous problem instances. The second method simplifies the energy function by `reducing' the number of unknown variables, and can also be used to generate a good initialization for the dynamic α-expansion algorithm by `reusing' dual variables.

We test the performance of our methods on energy functions encountered in the problems of stereo matching, and colour and object based segmentation. Experimental results show that our methods achieve a substantial improvement in the performance of α-expansion, as well as other popular algorithms such as sequential tree-reweighted message passing, and max-product belief propagation. In most cases we achieve a 10-15 times speed-up in the computation time. Our modified α-expansion algorithm provides similar performance to Fast-PD. However, it is much simpler and can be made orders of magnitude faster by using the initialization schemes proposed in the paper.

63. Integrated Feature Selection and Higher-order Spatial Feature Extraction for Object Categorization

David Liu, Gang Hua, Paul Viola, Tsuhan Chen

In computer vision, the bag-of-visual words image representation has been shown to yield good results. Recent work has shown that modeling the spatial relationship between visual words further improves performance. Previous work extracts higher-order spatial features exhaustively. However, these spatial features are expensive to compute. We propose a novel method that simultaneously performs feature selection and feature extraction. Higher-order spatial features are progressively extracted based on selected lower order ones, thereby avoiding exhaustive computation. The method can be based on any additive feature selection algorithm such as boosting. Experimental results show that the method is computationally much more efficient than previous approaches, without sacrificing accuracy.

64. Dimensionality Reduction using Covariance Operator Inverse Regression

Minyoung Kim, Vladimir Pavlovic

We consider the task of dimensionality reduction for regression (DRR) whose goal is to find a low dimensional representation of input covariates, while preserving the statistical correlation with output targets. DRR is particularly suited for visualization of high dimensional data as well as the efficient regressor design with a reduced input dimension. In this paper we propose a novel nonlinear method for DRR that exploits the kernel Gram matrices of input and output. While most existing DRR techniques rely on the inverse regression, our approach removes the need for explicit slicing of the output space using covariance operators in RKHS. This unique property make DRR applicable to problem domains with high dimensional output data with potentially significant amounts of noise. Although recent kernel dimensionality reduction algorithms make use of RKHS covariance operators to quantify conditional dependency between the input and the targets via the dimension-reduced input, they are either limited to a transduction setting or linear input subspaces and restricted to non-closed-form solutions. In contrast, our approach provides a closed-form solution to the nonlinear basis functions on which any new input point can be easily projected. We demonstrate the benefits of the proposed method in a comprehensive set of evaluations on several important regression problems that arise in computer vision.

65. Regression from Patch-kernel

Shuicheng Yan, Xi Zhou, mark John, Ming Liu, Thomas Huang

In this paper, we present a patch-based regression framework for addressing the human age and head pose estimation problems. Firstly, each image is encoded as an ensemble of orderless *coordinate patches*, the global distribution of which is described by Gaussian Mixture Models (GMM), and then each image is further expressed as a specific distribution model by Maximum a Posteriori adaptation from the global GMM. Then the *patch-kernel* is designed for characterizing the Kullback-Leibler divergence between the derived models for any two images, and its discriminating power is further enhanced by a weak learning process, called *inter-modality similarity synchronization*. Finally, *kernel regression* is employed for ultimate human age or head pose estimation. These three stages are complementary to each other, and jointly minimize the regression error. The effectiveness of this regression framework is validated by three experiments: 1) on the YAMAHA aging database, our solution brings a more than 50% reduction in age estimation error compared with the best reported results; 2) on the FG-NET aging database, our solution based on raw image features performs even better than the state-of-the-art algorithms which require fine face alignment for extracting warped appearance features; and 3) on the CHIL head pose database, our solution significantly outperforms the best one reported in the CLEAR07 evaluation.

66. Large Margin Pursuit for a Conic Section Classifier

Santhosh Kodipaka, Arunava Banerjee, Baba C. Vemuri

Learning a discriminant becomes substantially more difficult when the datasets are high-dimensional and the available samples are few. This is often the case in computer vision and medical diagnosis applications. A novel Conic Section classifier (CSC) was recently introduced in the literature to handle such datasets, wherein each class was represented by a conic section parameterized by its focus, directrix and eccentricity. The discriminant boundary was the locus of all points that are equi-eccentric relative to each class-representative conic section. Simpler boundaries were preferred for the sake of generalizability.

In this paper, we improve the performance of the two-class classifier via a large margin pursuit. When formulated as a non-linear optimization problem, the margin computation is demonstrated to be hard, especially due to the high dimensionality of the data. Instead, we present a geometric algorithm to compute the distance of a point to the non-linear discriminant boundary generated by the CSC in the input space. We then introduce a large margin pursuit in the learning phase so as to enhance the generalization capacity of the classifier. We validate the algorithm on real datasets and show favorable classification rates in comparison to many existing state-of-the-art binary classifiers as well as the CSC without margin pursuit.

67. Robust learning of discriminative projection for multicategory classification on the Stiefel manifold

Duc-Son Pham, Svetha Venkatesh

Learning a robust projection with a small number of training samples is still a challenging problem in face recognition, especially when the unseen faces have extreme variation in pose, illumination, and facial expression. To address this problem, we propose a framework formulated under statistical learning theory that facilitates robust learning of a discriminative projection. Dimensionality reduction using the projection matrix is combined with a linear classifier in the regularized framework of lasso regression. The projection matrix in conjunction with the classifier parameters are then found by solving an optimization problem over the Stiefel manifold. The experimental results on standard face databases suggest that the proposed method outperforms some recent regularized techniques when the number of training samples is small.

68. Joint learning and dictionary construction for pattern recognition

Duc-Son Pham, Svetha Venkatesh

We propose a joint representation and classification framework that achieves the dual goal of finding the most discriminative sparse overcomplete encoding and optimal classifier parameters. Formulating an optimization problem that combines the objective function of the classification with the representation error of both labeled and unlabeled data, constrained by sparsity, we propose an algorithm that alternates between solving for subsets of parameters, whilst preserving the sparsity. The method is then evaluated over two important classification problems in computer vision: object categorization of natural images using the Caltech 101 database and face recognition using the Extended Yale B face database. The results show that the proposed method is competitive against other recently proposed sparse overcomplete counterparts and considerably outperforms many recently proposed face recognition techniques when the number training samples is small.

69. Similarity-based cross-layered hierarchical representation for object categorization

Sanja Fidler, Marko Boben, Aleš Leonardis

This paper proposes a new concept in hierarchical representations that exploits features of different granularity and specificity coming from all layers of the hierarchy. The concept is realized within a cross-layered compositional representation learned from the visual data. We show how similarity connections among discrete labels within and across hierarchical layers can be established in order to produce a set of layer-independent shape-terminals, i.e., shapinals. We thus break the traditional notion of hierarchies and show how the category-specific layers can make use of all the necessary features stemming from all hierarchical layers. This, on the one hand, brings higher generalization into the representation, yet on the other hand, it also encodes the notion of scales directly into the hierarchy, thus enabling a multi-scale representation of object categories. By focusing on shape information only, the approach is tested on the Caltech 101 dataset demonstrating good performance in comparison with other state-of-the-art methods.

70. Learning and Using Taxonomies For Fast Visual Category Recognition

Gregory Griffin, Pietro Perona

The computational complexity of current visual categorization algorithms scales linearly at best with the number of categories. The goal of classifying simultaneously $\underline{N}_{\text{cat}} = 10^4 - 10^5$ visual categories requires sub-linear classification costs. We explore algorithms for automatically building classification trees which have, in principle, log N_{cat} complexity. We find that a greedy algorithm that recursively splits the set of categories into the two minimally confused subsets achieves 5-20 fold speedups at a small cost in classification performance. Our approach is independent of the specific classification algorithm used. A welcome byproduct of our algorithm is a very reasonable taxonomy of the Caltech-256 dataset.

10:30am – 12:15pm Oral Session O1A-1: Color, Illumination and Reflectance (La Perouse)

1. What do color changes reveal about an outdoor scene?

Kalyan Sunkavalli, Fabiano Romeiro, Wojciech Matusik, Todd Zickler, Hanspeter Pfister

In an extended image sequence of an outdoor scene, one observes changes in color induced by variations in the spectral composition of daylight. This paper proposes a model for these temporal color changes and explores its use for the analysis of outdoor scenes from time-lapse video data. We show that the time-varying changes in direct sunlight and ambient skylight can be recovered with this model, and that an image sequence can be decomposed into two corresponding components. The decomposition provides access to both radiometric and geometric information about a scene, and we demonstrate how this can be exploited for a variety of visual tasks, including color-constancy, background subtraction, shadow detection, scene reconstruction, and camera geo-location.

2. A Theory of Defocus via Fourier Analysis

Paolo Favaro, Alessandro Duci

In this paper we present a novel theory to analyze defocused images of a volume density by exploiting well-known results in Fourier analysis and the singular value decomposition. This analysis is fundamental in two respects: First, it gives a deep insight into the basic mechanisms of image formation of defocused images, and second, it shows how to incorporate additional a-priori knowledge about the geometry and photometry of the scene in restoration algorithms. For instance, we show that the case of a scene made of a single surface results in a simple constraint in the Fourier domain. We derive two basic types of algorithms for volumetric reconstruction: One based on a dense set of defocused images, and one based on a sparse set of defocused images. While the first one excels in simplicity, the second one is of more practical use. Both algorithms are tested on real and synthetic data.

3. Single-Image Vignetting Correction Using Radial Gradient Symmetry

Yuanjie Zheng, Jingyi Yu, Sing-Bing Kang, Stephen Lin, Chandra Kambhamettu

In this paper, we present a novel single-image vignetting method based on the symmetric distribution of the radial gradient (RG). The radial gradient is the image gradient along the radial direction with respect to the image center. We show that the RG distribution for natural images without vignetting is generally symmetric. However, this distribution is skewed by vignetting. We develop two variants of this technique, both of which remove vignetting by minimizing asymmetry of the RG distribution. Compared with prior approaches to single-image vignetting correction, our method does not require segmentation and the results are generally better. Experiments show our technique works for a wide range of images and it achieves a speed-up of 4-5 times compared with a state-of-the-art method.

4. Least Squares Surface Reconstruction from Measured Gradient Fields

Matthew Harker, Paul O'Leary

This paper presents a new method for the reconstruction of a surface from its x and y gradient field, measured, for example, via Photometric Stereo. The new algorithm produces the unique discrete surface whose gradients are equal to the measured gradients in the global vertical-distance least-squares sense. We show that it has been erroneously believed that this problem has been solved before via the solution of a Poisson equation. The numerical behaviour of the algorithm allows for reliable surface reconstruction on exceedingly large scales, e.g., full digital images; moreover, the algorithm is direct, i.e., non-iterative. We demonstrate the algorithm with synthetic data as well as real data obtained via photometric stereo. The algorithm does not exhibit a low-frequency bias and is not unrealistically constrained to arbitrary boundary conditions as in previous solutions. In fact, it is the first algorithm which can reconstruct a surface of polynomial degree two or higher exactly. It is hence the first viable algorithm for online industrial inspection where real defects (as opposed to phantom defects) must be identified in a robust manner.

5. Intensity Statistics-based HSI Diffusion for Color Photo Denoising

Lei He, Chunming Li, Chenyang Xu

This paper presents a new image denoising model for real color photo noise removal. Our model is implemented in the hue, saturation and intensity (HSI) space. The hue and saturation denoising are combined and implemented as a complex total variation (TV) diffusion. The intensity denoising is based on a diffusion flow to minimize a new energy functional, which is constructed with intensity component statistics. Besides the common gradient-based edge stopping functions for anisotropic diffusion, specifically for color photo denoising, we incorporate an intensity-based brightness adjusting term in the new energy, which corresponds to the noise disturbance with respect to photo intensity. In addition, we use the gradient vector flow (GVF) as the new diffusion directions for more accurate and robust denoising. Compared with previous diffusion flows only based on regular image gradients, this model provides more accurate image structure and intensity noise characterization for better denoising. Comprehensive quantitative and qualitative experiments on color photos demonstrate the improved performance of the proposed model when compared with 14 recognized approaches and 2 commercial software.

10:30am – 12:15pm Oral Session O1A-2: Segmentation and Grouping (Cook)

1. Edge preserving spatially varying mixtures for image segmentation

Giorgos Sfikas, Christophoros Nikou, Nikolaos Galatsanos

A new hierarchical Bayesian model is proposed for image segmentation based on Gaussian mixture models (GMM) with a prior enforcing spatial smoothness. According to this prior, the local differences of the contextual mixing proportions (i.e., the probabilities of class labels) are Student's t-distributed. The generative properties of the Student's t-pdf allow this prior to impose smoothness and simultaneously model the edges between the segments of the image. A maximum a posteriori (MAP) expectation-maximization (EM) based algorithm is used for Bayesian inference. An important feature of this algorithm is that all the parameters are automatically estimated from the data in closed form. Numerical experiments are presented that demonstrate the superiority of the proposed model for image segmentation as compared to standard GMM-based approaches and to GMM segmentation techniques with “standard” spatial smoothness constraints.

2. Robust Higher Order Potentials for Enforcing Label Consistency

Pushmeet Kohli, Lubor Ladicky, Philip H. S. Torr

This paper proposes a novel framework for labelling problems which is able to combine multiple segmentations in a principled manner. Our method is based on higher order conditional random fields and uses potentials defined on sets of pixels (image segments) generated using unsupervised segmentation algorithms. These potentials enforce label consistency in image regions and can be seen as a strict generalization of the commonly used pairwise contrast sensitive smoothness potentials. The higher order potential functions used in our framework take the form of the Robust P^n model. This enables the use of powerful graph cut based move making algorithms for performing inference in the framework. We test our method on the problem of multi-class object segmentation by augmenting the conventional CRF used for object segmentation with higher order potentials defined on image regions. Experiments on challenging data sets show that integration of higher order potentials quantitatively and qualitatively improves results leading to much better definition of object boundaries. We believe that this method can be used to yield similar improvements for many other labelling problems.

3. Multi-label Image Segmentation via Point-wise Repetition

Gang Zeng, Luc Van Gool

Bottom-up segmentation tends to rely on local features. Yet, many natural and man-made objects contain repeating elements. Such structural and more spread-out features are important cues for segmentation but are more difficult to exploit. The difficulty also comes from the fact that repetition need not be perfect, and will actually rather be partial, approximate, or both in most cases. This paper presents a multi-label image segmentation algorithm that processes a single input image and efficiently discovers and exploits repeating elements without any prior knowledge about their shape, color or structure. The algorithm spells out the interplay between segmentation and repetition detection.

The key of our approach is a novel, point-wise concept of repetition. This is defined by point-wise mutual information and locally compares certain neighborhoods to accumulate evidence. This point-wise repetition measure naturally handles imperfect repetitions, and the parts with inconsistent appearances are recognized and assigned with low scores. An energy functional is proposed to include the point-wise repetition into the image segmentation process, which takes the form of a graph-cut minimization. Real scene images demonstrate the ability of our algorithm to handle partial and approximate repetition.

4. Segmentation by transduction

Olivier Duchenne, Jean-Yves Audibert, Renaud Keriven, Jean Ponce, Florent Segonne

This paper addresses the problem of segmenting an image into regions consistent with user-supplied seeds (e.g., a sparse set of broad brush strokes). We view this task as a statistical transductive inference, in which some pixels are already associated with given zones and the remaining ones need to be classified. Our method relies on the Laplacian graph regularizer, a powerful manifold learning tool that is based on the estimation of variants of the Laplace-Beltrami operator and is tightly related to diffusion processes. Segmentation is modeled as the task of finding matting coefficients for unclassified pixels given known matting coefficients for seed pixels. The proposed algorithm essentially relies on a high margin assumption in the space of pixel characteristics. It is simple, fast, and accurate, as demonstrated by qualitative results on natural images and a quantitative comparison with state-of-the-art methods on the Microsoft GrabCut segmentation database.

5. Using Contours to Detect and Localize Junctions in Natural Images

Michael Maire, Pablo Arbeláez, Charless Fowlkes, Jitendra Malik

Contours and junctions are important cues for perceptual organization and shape recognition. Detecting junctions locally has proved problematic because the image intensity surface is confusing in the neighborhood of a junction. Edge detectors also do not perform well near junctions. Current leading approaches to junction detection, such as the Harris operator, are based on 2D variation in the intensity signal. However, a drawback of this strategy is that it confuses textured regions with junctions. We believe that the right approach to junction detection should take advantage of the contours that are incident at a junction; contours themselves can be detected by processes that use more global approaches. In this paper, we develop a new high-performance contour detector using a combination of local and global cues. This contour detector provides the best performance to date (F=0.70) on the Berkeley Segmentation Dataset (BSDS) benchmark. From the resulting contours, we detect and localize candidate junctions, taking into account both contour salience and geometric configuration. We show that improvements in our contour model lead to better junctions. Our contour and junction detectors both provide state of the art performance.

1:45pm – 3:45pm Poster Session P1P-1: Segmentation and Grouping (Summit Hall)

1. Interactive Image Matting for Multiple Layers

Dheeraj Singaraju, René Vidal

Image matting deals with finding the probability that each pixel in an image belongs to a user specified `object' or to the remaining `background'. Most existing methods estimate the mattes for two groups only. Moreover, most of these methods estimate the mattes with a particular bias towards the object and hence the resulting mattes do not sum up to 1 across the different groups. In this work, we propose a general framework to estimate the alpha mattes for multiple image layers. The mattes are estimated as the solution to the Dirichlet problem on a combinatorial graph with boundary conditions. We consider the constrained optimization problem that enforces the alpha mattes to take values in [0, 1] and sum up to 1 at each pixel. We also analyze the properties of the solution obtained by relaxing either of the two constraints. Experiments demonstrate that our proposed method can be used to extract accurate mattes of multiple objects with little user interaction.

2. Clustering and Dimensionality Reduction on Riemannian Manifolds

Alvina Goh, René Vidal

We propose a novel algorithm for clustering data sampled from multiple submanifolds of a Riemannian manifold. First, we learn a representation of the data using generalizations of local nonlinear dimensionality reduction algorithms from Euclidean to Riemannian spaces. Such generalizations exploit geometric properties of the Riemannian space, particularly its Riemannian metric. Then, assuming that the data points from different groups are separated, we show that the null space of a matrix built from the local representation gives the segmentation of the data. Our method is computationally simple and performs automatic segmentation without requiring user initialization. We present results on 2-D motion segmentation and diffusion tensor imaging segmentation.

3. Accurate Polyp Segmentation for 3D CT Colongraphy Using Multi-Staged Probabilistic Binary Learning and Compositional Model

Le Lu, Adrian Barbu, Matthias Wolf, Jianming Liang, Marcos Salganicoff, Dorin Comaniciu

Accurate and automatic colonic polyp segmentation and measurement in Computed Tomography (CT) has significant importance for 3D polyp detection, classification, and more generally computer aided diagnosis of colon cancers. In this paper, we propose a three-staged probabilistic binary classification approach for automatically segmenting polyp voxels from their surrounding tissues in CT. Our system integrates low-, and mid-level information for discriminative learning under local polar coordinates which align on the 3D colon surface around detected polyp. More importantly, our supervised learning system has flexible modeling capacity, which offers a principled means of encoding semantic, clinical expert annotations of colonic polyp tissue identification and segmentation. The learning generality to unseen data is bounded by boosting and stacked generality. Extensive experimental results on polyp segmentation performance evaluation and robustness testing with disturbances (using both training data and unseen data) are provided to validate our presented approach. The reliability of polyp segmentation and measurement has been largely increased to 98.2% (i.e., errors $\leq$ 3 mm), compared with other state of art work of about 75% ~ 80%.

4. Unsupervised Learning of Finite Mixtures Using Entropy Regularization and Its Application to Image Segmentation

Zhiwu Lu, Yuxin Peng, Jianguo Xiao

When fitting finite mixtures to multivariate data, it is crucial to select the appropriate number of components. Under regularization theory, we aim to resolve this “unsupervised” learning problem via regularizing the likelihood by the full entropy of posterior probabilities for finite mixture fitting. Two deterministic annealing implementations are further proposed for this entropy regularized likelihood (ERL) learning. Through some asymptotic analysis of the deterministic annealing ERL (DAERL) learning, we find that the global minimization of the ERL function in an annealing way can lead to automatic model selection on finite mixtures and also make our DAERL algorithms less sensitive to initialization than the standard EM algorithm. The simulation experiments then demonstrate that our algorithms can provide some promising results just as our theoretic analysis. Moreover, our algorithms are evaluated in the application of unsupervised image segmentation and shown to outperform other state-of-the-art methods.

5. Symmetric Multi-View Stereo Reconstruction From Planar Camera Arrays

Matthieu Maitre, Yoshihisa Shinagawa, Minh N. Do

We present a novel stereo algorithm which performs surface reconstruction from planar camera arrays. It incorporates the merits of both generic camera arrays and rectified binocular setups, recovering large surfaces like the former and performing efficient computations like the latter. First, we introduce a rectification algorithm which gives freedom in the design of camera arrays and simplifies photometric and geometric computations. We then define a novel set of data-fusion functions over 4-neighborhoods of cameras, which treat all cameras symmetrically and enable standard binocular stereo algorithms to handle arrays with arbitrary number of cameras. In particular, we introduce a photometric fusion function which handles partial visibility and extracts depth information along both horizontal and vertical baselines. Finally, we show that layered depth images and sprites with depth can be efficiently extracted from the rectified 3D space. Experimental results on real images confirm the effectiveness of the proposed method, which reconstructs dense surfaces larger by 20% on Tsukuba.

6. Video Segmentation: Propagation, Validation and Aggregation of a Preceding Graph

Siying Liu, Guo Dong, Chye Hwang Yan, Sim Heng Ong

In this work, video segmentation is viewed as an efficient intra-frame grouping temporally reinforced by a strong inter-frame coherence. Traditional approaches simply regard pixel motions as another prior in the MRF-MAP framework. Since pixel pre-grouping is inefficiently performed on every frame, the strong correlation between inter-frame groupings is largely underutilized. We exploit the inter-frame correlation to propagate trustworthy groupings from the previous frame. A preceding graph is constructed and labeled for the previous frame. It is temporally propagated to the current frame and validated by similarity measures. All unlabeled subgraphs are spatially aggregated for the final grouping. Experimental results show that the proposed approach is highly efficient for spatio-temporal segmentation. It makes good use of temporal correlation and produces satisfactory grouping results.

7. Boundary Snapping for Robust Image Cutouts

Eyal Zadicario, Shai Avidan, Alon Shmueli, Daniel Cohen-Or

Boundary Snapping is an interactive image cutout algorithm that requires a small number of user supplied control points, or landmarks, to infer the cutout contour. The key idea is to match the appearance of all points along the desired contour to the landmark points, where appearance is given by an intensity profile perpendicular to the boundary. An optimization process attempts to find a contour that maximizes the similarity score of its points with the landmarks. This approach works well in the typical case where the foreground and background differ in appearance, as well as in challenging cases where the subject is clearly perceived, but the regions on both sides of the boundary are similar and cannot be easily discriminated. By enabling the user to define the boundary points directly, the technique is not limited to boundaries that necessarily have to be the most salient or high gradient feature in the region. It can also be used for margin cutout around the boundary. The use of multiple control points along the boundary can handle spatially varying attributes as both foreground and background may change in appearance along the boundary. The final result is accurate, because it allows the user to enforce hard constraints on the boundary directly, at the expense of moderate user labor in positioning the landmark points. Finally, the algorithm is fast, works on a variety of images, and handles situations where the boundary is not obvious.

8. Tracking Distributions with an Overlap Prior

Ismail Ben Ayed, Shuo Li, Ian Ross

Recent studies have shown that embedding similarity/dissimilarity measures between distributions in the variational level set framework can lead to effective object segmentation/tracking algorithms. In this connection, existing methods assume implicitly that the overlap between the distributions of image data within the object and its background has to be minimal. Unfortunately, such assumption may not be valid in many important applications. This study investigates an overlap prior, which embeds knowledge about the overlap between the distributions of the object and the background in level set tracking. It consists of evolving a curve to delineate the target object in the current frame. The level set curve evolution equation is sought following the maximization of a functional containing three terms: (1) an original overlap prior which measures the conformity of overlap between the nonparametric (kernel-based) distributions within the object and the background to a learned description, (2) a term which measures the similarity between a model distribution of the object and the sample distribution inside the curve, and (3) a regularization term for smooth segmentation boundaries. The Bhattacharyya coefficient is used as an overlap measure. Apart from leading to a method which is more versatile than current ones, the overlap prior speeds up significantly the curve evolution. Comparisons and results demonstrate the advantages of the proposed prior over related methods, and its usefulness in important applications such as the left ventricle tracking in Magnetic Resonance (MR) images.

9. Globally Optimal Surface Segmentation Using Regional Properties of Segmented Objects

Xin Dou, Wu Xiaodong, Andreas Wahle, Milan Sonka

Efficient segmentation of globally optimal surfaces in volumetric images is a central problem in many medical image analysis applications. Intra-class variance has been successfully utilized, for instance, in the Chan-Vese model especially for images without prominent edges. In this paper, we study the optimization problem of detecting a region (volume) between two coupled smooth surfaces by minimizing the intra-class variance using an efficient polynomial-time algorithm. Our algorithm is based on the shape probing technique in computational geometry and computes a sequence of minimum-cost closed sets in a derived parametric graph. The method has been validated on computer-synthetic volumetric images and in X-ray CT-scanned datasets of plexiglas tubes of known sizes. Its applicability to clinical data sets was demonstrated in human CT image data. The achieved results were highly accurate with mean signed surface positioning errors of the inner and outer walls of the tubes of +0.013mm and 0.012mm, respectively, given a voxel size of $0.39 \times 0.39 \times 0.6\text{mm}^3$. Comparing with the original Chan-Vese method, our algorithm expressed higher robustness. With its polynomial-time efficiency, our algorithm is ready to be extended to higher-dimensional image segmentation. In addition, the developed technique is of its own interest. We expect that it can shed some light on solving other important optimization problems arising in computer vision. To the best of our knowledge, the shape probing technique is for the first time introduced into the field of computer vision.

10. A Bayesian Approach for Image Segmentation with Shape Priors

Hang Chang, Qing Yang, Bahram Parvin

Color and texture have been widely used in image segmentation; however, their performance is often hindered by scene ambiguities, overlapping objects, or missing parts. In this paper, we propose an interactive image segmentation approach with shape prior models within a Bayesian framework. Interactive features, through mouse strokes, reduce ambiguities, and the incorporation of shape priors enhances quality of the segmentation where color and/or texture are not solely adequate. The novelties of our approach are in (i) formulating the segmentation problem in a well-defined Bayesian framework with multiple shape priors, (ii) efficiently estimating parameters of the Bayesian model, and (iii) multi-object segmentation through user-specified priors. We demonstrate the effectiveness of our method on a set of natural and synthetic images.

11. Selective Hidden Random Fields: Exploiting Domain Specific Saliency for Event Classification

Vidit Jain, Amit Singhal, Jiebo Luo

Classifying an event captured in an image is useful for understanding the contents of the image. The captured event provides context to refine models for the presence and appearance of various entities, such as people and objects, in the captured scene. Such contextual processing facilitates the generation of better abstractions and annotations for the image. Consider a typical set of consumer images with sports-related content. These images are taken mostly by amateur photographers, and often at a distance. In the absence of manual annotation or other sources of information such as time and location, typical recognition tasks are formidable on these images. Identifying the sporting event in these images provides a context for further recognition and annotation tasks. We propose to use the domainspecific saliency of the appearances of the playing surfaces, and ignore the noninformative parts of the image such as crowd regions, to discriminate among different sports. To this end, we present a variation of the hidden-state conditional random field that selects a subset of the observed features suitable for classification. The inferred hidden variables in this model represent a selection criteria desirable for the problem domain. For sports-related images, this selection criteria corresponds to the segmentation of the playing surface in the image. We demonstrate the utility of this model on consumer images collected from the Internet.

12. Learning Class-Specific Affinities for Image Labelling

Dhruv Batra, Rahul Sukthankar, Tsuhan Chen

Spectral clustering and eigenvector-based methods have become increasingly popular in segmentation and recognition. Although the choice of the pairwise similarity metric (or affinities) greatly influences the quality of the results, this choice is typically specified outside the learning framework. In this paper, we present an algorithm to learn class-specific similarity functions. Mapping our problem in a Conditional Random Fields (CRF) framework enables us to pose the task of learning affinities as parameter learning in undirected graphical models. There are two significant advances over previous work. First, we learn the affinity between a pair of data-points as a function of a pairwise feature and (in contrast with previous approaches) the classes to which these two data-points were mapped, allowing us to work with a richer class of affinities. Second, our formulation provides a principled probabilistic interpretation for learning all of the parameters that define these affinities. Using ground truth segmentations and labellings for training, we learn the parameters with the greatest discriminative power (in an MLE sense) on the training data. We demonstrate the power of this learning algorithm in the setting of joint segmentation and recognition of object classes. Specifically, even with very simple appearance features, the proposed method achieves state-of-the-art performance on standard datasets.

13. Segmentation of Left Ventricle From 3D Cardiac MR Sequences Using A Subject-Specific Dynamical Model

Yun Zhu, Xenophon Papademetris, Albert Sinusas, James Duncan

Statistical model-based segmentation of the left ventricle from cardiac images has received considerable attention in recent years. While a variety of statistical models have been shown to improve segmentation results, most of them are either static models (SM) which neglect the temporal coherence of a cardiac sequence or generic dynamical models (GDM) which neglect the inter-subject variability of cardiac shapes and deformations. In this paper, we use a subject-specific dynamical model (SSDM) that handles inter-subject variability and temporal dynamics (intra-subject variability) simultaneously. It can progressively identify the specific motion patterns of a new cardiac sequence based on the segmentations observed in the past frames. We formulate the integration of the SSDM into the segmentation process in a recursive Bayesian framework in order to segment each frame based on the intensity information from the current frame and the prediction from the past frames. We perform "Leave-one-out" test on 32 sequences to validate our approach. Quantitative analysis of experimental results shows that the segmentation with the SSDM outperforms those with the SM and GDM by having better global and local consistencies with the manual segmentation.

14. Geo-spatial Aerial Video Processing for Scene Understanding and Object Tracking

Jiangjian Xiao, Cheng Hui, Feng Han, Harpreet Sawhney

This paper presents an approach to extracting and using semantic layers from low altitude aerial videos for scene understanding and object tracking. The input video is captured by low flying aerial platforms and typically consists of strong parallax from non-ground-plane structures. A key aspect of our approach is the use of geo-registration of video frames to reference image databases (such as those available from Terraserver and Google satellite imagery) to establish a geo-spatial coordinate system for pixels in the video. Geo-registration enables Euclidean 3D reconstruction with absolute scale unlike traditional monocular structure from motion where continuous scale estimation over long periods of time is an issue. Geo-registration also enables correlation of video data to other stored information sources such as GIS (Geo-spatial Information System) databases. In addition to the geo-registration and 3D reconstruction aspects, the key contributions of this paper include: (1) exploiting appearance and 3D shape constraints derived from geo-registered videos for labeling of structures such as buildings, foliage, and roads for scene understanding, and (2) elimination of moving object detection and tracking errors using 3D parallax constraints and semantic labels derived from geo-registered videos. Experimental results on extended time aerial video data demonstrates the qualitative and quantitative aspects of our work.

15. Generalised Blurring Mean-Shift Algorithms for Nonparametric Clustering

Miguel Á. Carreira-Perpiñán

Gaussian blurring mean-shift (GBMS) is a nonparametric clustering algorithm, having a single bandwidth parameter that controls the number of clusters. The algorithm iteratively shrinks the data set under the application of a mean-shift update, stops in just a few iterations and yields excellent clusterings. We propose several families of generalised GBMS (GGBMS) algorithms based on explicit, implicit and exponential updates, and depending on a step-size parameter. We give conditions on the step size for the convergence of these algorithms and show that the convergence rate for Gaussian clusters ranges from sublinear to linear, cubic and even higher order depending on the update and step size. We show that the algorithms are related to spectral clustering if using a random-walk matrix with modified eigenvalues and updated after each iteration, and show the relation with methods developed for surface smoothing in the computer graphics literature. Detailed experiments in toy problems and image segmentation show that, while all the GGBMS algorithms can achieve essentially the same result (for appropriate settings of the bandwidth and step size), they significantly differ in runtime, with slightly overrelaxed explicit updates being fastest in practice.

16. Auto-Context and Its Application to High-level Vision Tasks

Zhuowen Tu

The notion of using context information for solving high-level vision problems has been increasingly realized in the field. However, how to learn an effective and efficient context model, together with the image appearance, remains mostly unknown. The current literature using Markov Random Fields (MRFs) and Conditional Random Fields (CRFs) often involves specific algorithm design, in which the modeling and computing stages are studied in isolation. In this paper, we propose an **auto-context** algorithm. Given a set of training images and their corresponding label maps, we first learn a classifier on local image patches. The discriminative probability (or classification confidence) maps by the learned classifier are then used as context information, in addition to the original image patches, to train a new classifier. The algorithm then iterates to approach the ground truth. Auto-context learns an integrated low-level and context model, and is very general and easy to implement. Under nearly the identical parameter setting in the training, we apply the algorithm on three challenging vision applications: object segmentation, human body configuration, and scene region labeling. It typically takes about 30 ~ 70 seconds to run the algorithm in testing. Moreover, the scope of the proposed algorithm goes beyond high-level vision. It has the potential to be used for a wide variety of problems of multi-variate labeling.

17. Motion Segmentation via Robust Subspace Separation in the Presence of Outlying, Incomplete, or Corrupted Trajectories

Shankar Rao, Roberto Tron, René Vidal, Yi Ma

We examine the problem of segmenting tracked feature point trajectories of multiple moving objects in an image sequence. Using the affine camera model, this motion segmentation problem can be cast as the problem of segmenting samples drawn from a union of linear subspaces. Due to limitations of the tracker, occlusions and the presence of nonrigid objects in the scene, the obtained motion trajectories may contain grossly mistracked features, missing entries, or not correspond to any valid motion model. In this paper, we develop a robust subspace separation scheme that can deal with all of these practical issues in a unified framework. Our methods draw strong connections between lossy compression, rank minimization, and sparse representation. We test our methods extensively and compare their performance to several extant methods with experiments on the Hopkins 155 database. Our results are on par with state-of-the-art results, and in many cases exceed them. All MATLAB code and segmentation results are publicly available for peer evaluation at http://perception.csl.uiuc.edu/coding/motion/.

18. Principled Fusion of High-level Model and Low-level Cues for Motion Segmentation

Arasanathan Thayananthan, Masahiro Iwasaki, Roberto Cipolla

High-level generative models provide elegant descriptions of videos and are commonly used as the inference framework in many unsupervised motion segmentation schemes. However, approximate inference in these models often require ad-hoc initialization to avoid local minima issues. Low-level cues, obtained independently from the high-level model, can constrain the search space and reduce the chance of inference algorithms falling into a local minima. This paper introduces a novel principled fusion framework where, local hierarchical superpixels segmentation of images are used to capture local motion. The low-level cues such as local motion, on their own, not adequate to obtain full motion segmentation as occlusion needs to be handled globally. We fuse the low-level motion cues with the high-level model in a principled manner to surmount the shortcomings of using only the high-level model or low-level cues to perform motion segmentation. The fused model contains both continuous and discrete variables which forms a number of Markov Random fields. Variational approximation or belief propagation algorithms cannot be applied due to the complex interactions between the variables. Hence, approximate inference is performed using expectation propagation (EP) algorithm. The scheme is demonstrated by performing motion segmentation in two video sequences.

19. Unsupervised estimation of segmentation quality using nonnegative factorization

Roman Sandler, Michael Lindenbaum

We propose an unsupervised method for evaluating image segmentation. Common methods are typically based on evaluating smoothness within segments and contrast between them, and the measure they provide is not explicitly related to segmentation errors. The proposed approach differs from these methods on several important points and has several advantages over them.

First, it provides a meaningful, quantitative assessment of segmentation quality, in precision/recall terms, which were applicable so far only for supervised evaluation. Second, it builds on a new image model, which characterizes the segments as a mixture of basic feature distributions. The precision/recall estimates are then obtained by a nonnegative matrix factorization (NMF) process. A third important advantage is that the estimates, which are based on intrinsic properties of the specific image being evaluated and not on a comparison to typical images (learning), are relatively robust to context factors such as image quality or the presence of texture.

Experimental results demonstrate the accuracy of the precision/recall estimates in comparison to ground truth based on human judgment. Moreover, it is shown that tuning a segmentation algorithm using the unsupervised measure improves the algorithm's quality (as measured by a supervised method).

20. Graph cut based image segmentation with connectivity priors

Sara Vicente, Vladimir Kolmogorov, Carsten Rother

Graph cut is a popular technique for interactive image segmentation. However, it has certain shortcomings. In particular, graph cut has problems with segmenting thin elongated objects due to the “shrinking bias”. To overcome this problem, we propose to impose an additional connectivity prior, which is a very natural assumption about objects. We formulate several versions of the connectivity constraint and show that the corresponding optimization problems are all NP-hard.

For some of these versions we propose two optimization algorithms: (i) a practical heuristic technique which we call DijkstraGC, and (ii) a slow method based on problem decomposition which provides a lower bound on the problem. We use the second technique to verify that for some practical examples DijkstraGC is able to find the global minimum.

21. High Resolution Matting via Interactive Trimap Segmentation

Christoph Rhemann, Carsten Rother, Alex Rav-Acha, Toby Sharp

We present a new approach to the matting problem which splits the task into two steps: interactive trimap extraction followed by trimap-based alpha matting. By doing so we gain considerably in terms of speed and quality and are able to deal with high resolution images. This paper has three contributions: (i) a new trimap segmentation method using parametric max-flow; (ii) an alpha matting technique for high resolution images with a new gradient preserving prior on alpha; (iii) a database of 27 ground truth alpha mattes of still objects, which is considerably larger than previous databases and also of higher quality. The database is used to train our system and to validate that both our trimap extraction and our matting method improve on state-of-the-art techniques.

22. Real-time 3D Segmentation of the Left Ventricle Using Deformable Subdivision Surfaces

Fredrik Orderud, Stein Inge Raben

In this paper, we extend a computationally efficient framework for real-time 3D tracking and segmentation to support deformable subdivision surfaces. Segmentation is performed in a sequential state-estimation fashion, using an extended Kalman filter to estimate shape and pose parameters for the subdivision surface. As an example, we have integrated Doo-Sabin subdivision surfaces into the framework. Furthermore, we provide a method for evaluating basis functions for Doo-Sabin surfaces at arbitrary parameter values. These basis functions are precomputed during initialization, and later used during segmentation to quickly evaluate surface points used for edge detection.

Fully automatic tracking and segmentation of the left ventricle is demonstrated in a dataset of 21 3D echocardiography recordings. Successful segmentation was achieved in all cases, with limits of agreement (mean±1.96SD) for point to surface distance of 2.2±0.8 mm compared to manually verified segmentations. Real-time segmentation at a rate of 25 frames per second consumed a CPU load of 8%.

23. Globally Optimal Shape-based Tracking in Real-time

Thomas Schoenemann, Daniel Cremers

Most algorithms for real-time tracking of deformable shapes provide sub-optimal solutions for a suitable energy minimization task: The search space is typically considered too large to allow for globally optimal solutions.

In this paper we show that -- under reasonable constraints on the object motion -- one can guarantee global optimality while maintaining real-time requirements. The problem is cast as finding the optimal cycle in a graph spanned by the prior template and the image. The underlying combinatorial algorithm is implemented on state-of-the-art graphics hardware. Solutions on FPGAs are conceivable.

Experimental results demonstrate long-term tracking of cars in real-time, while coping with challenging weather conditions. In particular, we show that the proposed tracking algorithm is highly robust to illumination changes and that it outperforms local tracking methods such as the level set method.

24. Matching Non-rigidly Deformable Shapes Across Images: A Globally Optimal Solution

Thomas Schoenemann, Daniel Cremers

While global methods for matching shapes to images have recently been proposed, so far research has focused on small deformations of a fixed template.

In this paper we present the first global method able to pixel-accurately match non-rigidly deformable shapes across images at amenable run-times. By finding cycles of optimal ratio in a four-dimensional graph -- spanned by the image, the prior shape and a set of rotation angles -- we simultaneously compute a segmentation of the image plane, a matching of points on the template to points on the segmenting boundary, and a decomposition of the template into a set of deformable parts.

In particular, the interpretation of the shape template as a collection of an a priori unknown number of deformable parts -- an important aspect of higher-level shape representations -- emerges as a byproduct of our matching algorithm. On real-world data of running people and walking animals, we demonstrate that the proposed method can match strongly deformed shapes, even in cases where simple shape measures and optic flow methods fail.

25. High Resolution Motion Layer Decomposition using Dual-space Graph Cuts

Thomas Schoenemann, Daniel Cremers

We introduce a novel energy minimization method to decompose a video into a set of super-resolved moving layers. The proposed energy corresponds to the cost of coding the sequence. It consists of a data term and two terms imposing regularity of the geometry and the intensity of each layer.

In contrast to existing motion layer methods, we perform graph cut optimization in the (dual) layer space to determine which layer is visible at which video position. In particular, we show how arising higher-order terms can be accounted for by a generalization of alpha expansions. Moreover, our model accurately captures long-term temporal consistency. To the best of our knowledge, this is the first work which aims at modeling details of the image formation process (such as camera blur and downsampling) in the context of motion layer decomposition. The experimental results demonstrate that energy minimization leads to a reconstruction of a video in terms of a superposition of multiple high-resolution motion layers.

26. Shape Priors in Variational Image Segmentation: Convexity, Lipschitz Continuity and Globally Optimal Solutions

Daniel Cremers, Frank R. Schmidt, Frank Barthel

In this work, we introduce a novel implicit representation of shape which is based on assigning to each pixel a probability that this pixel is inside the shape. This probabilistic representation of shape resolves two important drawbacks of alternative implicit shape representations such as the level set method: Firstly, the space of shapes is convex in the sense that arbitrary convex combinations of a set of shapes again correspond to a valid shape. Secondly, we prove that the introduction of shape priors into variational image segmentation leads to functionals which are convex with respect to shape deformations.

For a large class of commonly considered (spatially continuous) functionals, we prove that -- under mild regularity assumptions -- segmentation and tracking with statistical shape priors can be performed in a globally optimal manner. In experiments on tracking a walking person through a cluttered scene we demonstrate the advantage of global versus local optimality.

27. Kernel-based learning of cast shadows from a physical model of light sources and surfaces for low-level segmentation

Nicolas Martel-Brisson, Andre Zaccarin

In background subtraction, cast shadows induce silhouette distortions and object fusions hindering performance of high level algorithms in scene monitoring. We introduce a nonparametric framework to model surface behavior when shadows are cast on them. Based on physical properties of light sources and surfaces, we identify a direction in RGB space on which background surface values under cast shadows are found. We then model the posterior distribution of lighting attenuation under cast shadows and foreground objects, which allows differentiation of foreground and cast shadow values with similar chromaticity. The algorithms are completely unsupervised and take advantage of scene activity to learn model parameters. Spatial gradient information is also used to reinforce the learning process. Contributions are two-fold. Firstly, with a better model describing cast shadows on surfaces, we achieve a higher success rate in segmenting moving cast shadows in complex scenes. Secondly, obtaining such models is a step toward a full scene parametrization where light source properties, surface reflectance models and scene 3D geometry are estimated for low-level segmentation.

28. Partitioning of Image Datasets using Discriminative Context Information

Christoph H. Lampert

We propose a new method to partition an unlabeled dataset, called Discriminative Context Partitioning (DCP). It is motivated by the idea of splitting the dataset based only on how well the resulting parts can be separated from a context class of disjoint data points. This is in contrast to typical clustering techniques like K-means that are based on a generative model by implicitly or explicitly searching for modes in the distribution of samples. The discriminative criterion in DCP avoids the problems that density based methods have when the a priori assumption of multimodality is violated, when the number of samples becomes small in relation to the dimensionality of the feature space, or if the cluster sizes are strongly unbalanced. We formulate DCP's separation property as a large-margin criterion, and show how the resulting optimization problem can be solved efficiently. Experiments on the MNIST and USPS datasets of handwritten digits and on a subset of the Caltech256 dataset show that, given a suitable context, DCP can achieve good results even in situation where density-based clustering techniques fail.

29. Fast Texture Segmentation Using the Shape Operator and Active Contour

Nawal Houhou, Jean-Philippe Thiran, Xavier Bresson

We present an approach for unsupervised segmentation of natural and textural images based on active contour, differential geometry and information theoretical concept. More precisely, we propose a new texture descriptor which intrinsically defines the geometry of textural regions using the shape operator borrowed from differential geometry. Then, we use the popular Kullback-Leibler distance to define an active contour model which distinguishes the background and textural objects of interest represented by the probability density functions of our new texture descriptor. We prove the existence of a solution to the proposed segmentation model. Finally, a fast and easy to implement texture segmentation algorithm is introduced to extract meaningful objects. We present promising synthetic and real-world results and compare our algorithm to other state-of-the-art techniques.

30. Shape Prior Segmentation of Multiple Objects with Graph Cuts

Nhat Vu, B.S. Manjunath

We present a new shape prior segmentation method using graph cuts capable of segmenting multiple objects. The shape prior energy is based on a shape distance popular with level set approaches. We also present a multiphase graph cut framework to simultaneously segment multiple, possibly overlapping objects. The multiphase formulation differs from multiway cuts in that the former can account for object overlaps by allowing a pixel to have multiple labels. We then extend the shape prior energy to encompass multiple shape priors. Unlike variational methods, a major advantage of our approach is that the segmentation energy is minimized directly without having to compute its gradient, which can be a cumbersome task and often relies on approximations. Experiments demonstrate that our algorithm can cope with image noise and clutter, as well as partial occlusions and affine transformations of the shape.

31. Constrained Spectral Clustering through Affinity Propagation

Zhengdong Lu, Miguel Á. Carreira-Perpiñán

Pairwise constraints specify whether or not two samples should be in one cluster. Although it has been successful to incorporate them into traditional clustering methods, such as K-means, little progress has been made in combining them with spectral clustering. The major challenge in designing an effective constrained spectral clustering is a sensible combination of the scarce pairwise constraints with the original affinity matrix. We propose to combine the two sources of affinity by propagating the pairwise constraints information over the original affinity matrix. Our method has a Gaussian process interpretation and results in a closed-form expression for the new affinity matrix. Experiments show it outperforms state-of-the-art constrained clustering methods in getting good clusterings with fewer constraints, and yields good image segmentation with user-specified pairwise constraints.

32. Coherent Laplacian 3D Protrusion Segmentation

Fabio Cuzzolin, Diana Mateus, David Knossow, Edmond Boyer, Radu Horaud

In this paper, an analysis of locally linear embedding (LLE) in the context of clustering is developed. As LLE conserves the local affine coordinates of points, shape protrusions as high-curvature regions of the surface are preserved. Also, LLE's covariance constraint acts as a force stretching those protrusions and making them wider separated and lower dimensional. A novel scheme for unsupervised body-part segmentation along time sequences is thus proposed in which 3-D shapes are clustered after embedding. Clusters are propagated in time, and merged or split in an unsupervised fashion to accommodate changes of the body topology. Comparisons on synthetic, and real data with ground truth, are run with direct segmentation in 3-D by EM clustering and ISOMAP-based clustering. Robustness and the effects of topology transitions are discussed.

33. Combining Appearance Models and Markov Random Fields for Category Level Object Segmentation

Diane Larlus, Frédéric Jurie

Object models based on bag-of-words representations can achieve state-of-the-art performance for image classification and object localization tasks. However, as they consider objects as loose collections of local patches they fail to accurately locate object boundaries and are not able to produce accurate object segmentation. On the other hand, Markov Random Field models used for image segmentation focus on object boundaries but can hardly use the global constraints necessary to deal with object categories whose appearance may vary significantly. In this paper we combine the advantages of both approaches. First, a mechanism based on local regions allows object detection using visual word occurrences and produces a rough image segmentation. Then, a MRF component gives clean boundaries and enforces label consistency, guided by local image cues (color, texture and edge cues) and by long-distance dependencies. Gibbs sampling is used to infer the model. The proposed method successfully segments object categories with highly varying appearances in the presence of cluttered backgrounds and large view point changes. We show that it outperforms published results on the Pascal VOC 2007 dataset.

34. Normalized Tree Partitioning for Image Segmentation

Jingdong Wang, Yangqing Jia, Xian-Sheng Hua, Changshui Zhang, Long Quan

In this paper, we propose a novel graph based clustering approach with satisfactory clustering performance and low computational cost. It consists of two main steps: tree fitting and partitioning. We first introduce a probabilistic method to fit a tree to a data graph under the sense of minimum entropy. Then, we propose a novel tree partitioning method under a normalized cut criterion, called Normalized Tree Partitioning (NTP), in which a fast combinatorial algorithm is designed for exact bipartitioning. Moreover, we extend it to k-way tree partitioning by proposing an efficient best-first recursive bipartitioning scheme. Compared with spectral clustering, NTP produces the exact global optimal bipartition, introduces fewer approximations for k-way partitioning and can intrinsically produce superior performance. Compared with bottom-up aggregation methods, NTP adopts a global criterion and hence performs better. Last, experimental results on image segmentation demonstrate that our approach is more powerful compared with existing graph-based approaches.

35. FuzzyMatte: A Computationally Efficient Scheme for Interactive Matting

Yuanjie Zheng, Chandra Kambhamettu, Jingyi Yu, Bauer Thomas, Steiner Karl

In this paper, we propose an online interactive matting algorithm, which we call FuzzyMatte. Our framework is based on computing the fuzzy connectedness (FC) [20] from each unknown pixel to the known foreground and background. FC effectively captures the adjacency and similarity between image elements and can be efficiently computed using the strongest connected path searching algorithm. The final alpha value at each pixel can then be calculated from its FC. While many previous methods need to completely recompute the matte when new inputs are provided, FuzzyMatte effectively integrates these new inputs with the previously estimated matte by efficiently recomputing the FC value for a small subset of pixels. Thus, the computational overhead between each iteration of the refinement is significantly reduced. We demonstrate FuzzyMatte on a wide range of images. We show that FuzzyMatte updates the matte in an online interactive setting and generates high quality matte for complex images.

36. A Region Based Stereo Matching Algorithm Using Cooperative Optimization

Zengfu Wang, Zhigang Zheng

This paper presents a new stereo matching algorithm based on inter-regional cooperative optimization. The proposed algorithm uses regions as matching primitives and defines the corresponding region energy functional for matching by utilizing the color Statistics of regions and the constraints on smoothness and occlusion between adjacent regions. In order to obtain a more reasonable disparity map, a cooperative optimization procedure has been employed to minimize the matching costs of all regions by introducing the cooperative and competitive mechanism between regions. Firstly, a color based segmentation method is used to segment the reference image into regions with homogeneous color. Secondly, a local window-based matching method is used to determine the initial disparity estimate of each image pixel. And then, a voting based plane fitting technique is applied to obtain the parameters of disparity plane corresponding to each image region. Finally, the disparity plane parameters of all regions are iteratively optimized by an inter-regional cooperative optimization procedure until a reasonable disparity map is obtained. The experimental Results on Middlebury test set and real stereo images indicate that the performance of our method is competitive with the best stereo matching algorithms and the disparity maps recovered are close to the ground truth data.

37. Global Pose Estimation Using Non-Tree Models

Hao Jiang, David Martin

We propose a novel global pose estimation method to detect body parts of articulated objects in images based on non-tree graph models. There are two kinds of edges defined in the body part relation graph: Strong (tree) edges corresponding to the body plan that can enforce any type of constraint, and weak (non-tree) edges that express exclusion constraints arising from inter-part occlusion and symmetry conditions. We express optimal part localization as a multiple shortest path problem in a set of correlated trellises constructed from the graph model. Strong model edges generate the trellises, while weak model edges prohibit implausible poses by generating exclusion constraints among trellis nodes and edges. The optimization may be expressed as an integer linear program and solved using a novel two-stage relaxation scheme. Experiments show that the proposed method has a high chance of obtaining the globally optimal pose at low computational cost.

38. Building Segmentation for Densely Built Urban Regions Using Aerial LIDAR Data

Bogdan Matei, Harpreet Sawhney, Supun Samarasekera, Janet Kim, Rakesh Kumar

We present a novel building segmentation system for densely built areas, containing thousands of buildings per square kilometer. We employ solely sparse LIDAR (Light/Laser Detection & Ranging) 3D data, captured from an aerial platform, with resolution less than one point per square meter. The goal of our work is to create segmented and delineated buildings as well as structures on top of buildings without requiring scanning for the sides of buildings. Building segmentation is a critical component in many applications such as 3D visualization, robot navigation and cartography. LIDAR has emerged in recent years as a more robust alternative to 2D imagery because it acquires 3D structure directly, without the shortcomings of stereo in untextured regions and at depth discontinuities.

Our main technical contributions in this paper are: (i) a ground segmentation algorithm which can handle both rural regions, and heavily urbanized areas, where the ground is 20% or less of the data. (ii) a building segmentation technique, which is robust to buildings in close proximity to each other, sparse measurements and nearby structured vegetation clutter, and (iii) an algorithm for estimating the orientation of a boundary contour of a building, based on minimizing the number of vertices in a rectilinear approximation to the building outline, which can cope with significant quantization noise in the outline measurements.

We have applied the proposed building segmentation system to several urban regions with areas of hundreds of square kilometers each, obtaining average segmentation speeds of less than three minutes per km^2 on a standard Pentium processor. Extensive qualitative results obtained by overlaying the 3D segmented regions onto 2D imagery indicate accurate performance of our system.

39. Hybrid Body Representation for Integrated Pose Recognition, Localization and Segmentation

Cheng Chen, Guoliang Fan

We propose a hybrid body representation that represents each typical pose by both template-like view information and part-based structural information. Specifically, each body part as well as the whole body are represented by an off-line learned shape model where both region-based and edge-based priors are combined in a coupled shape representation. Part-based spatial priors are represented by a “star” graphical model. This hybrid body representation can synergistically integrate pose recognition, localization and segmentation into one computational flow. Moreover, as an important step for feature extraction and model inference, segmentation is involved in the low-level, mid-level and high-level vision stages, where top-down prior knowledge and bottom-up data processing is well integrated via the proposed hybrid body representation.

40. Image Segmentation via Convolution of a Level-Set Function with a Rigaut Kernel

Özlem N. Subakan, Baba C. Vemuri

Image segmentation is a fundamental task in Computer Vision and there are numerous algorithms that have been successfully applied in various domains. There are still plenty of challenges to be met with. In this paper, we consider one such challenge, that of achieving segmentation while preserving complicated and detailed features present in the image, be it a gray level or a textured image. We present a novel approach that does not make use of any prior information about the objects in the image being segmented. Segmentation is achieved using local orientation information, which is obtained via the application of a steerable Gabor filter bank, in a statistical framework. This information is used to construct a spatially varying kernel called the Rigaut Kernel, which is then convolved with the signed distance function of an evolving contour (placed in the image) to achieve segmentation. We present numerous experimental results on real images, including a quantitative evaluation. Superior performance of our technique is depicted via comparison to the state-of-the-art algorithms in literature.

41. Detecting and Matching Repeated Patterns for Automatic Geo-tagging in Urban Environments

Grant Schindler, Panchapagesan Krishnamurthy, Roberto Lublinerman, Yanxi Liu, Frank Dellaert

We present a novel method for automatically geo-tagging photographs of man-made environments via detection and matching of repeated patterns. Highly repetitive environments introduce numerous correspondence ambiguities and are problematic for traditional wide-baseline matching methods. Our method exploits the highly repetitive nature of urban environments, detecting multiple perspectively distorted periodic 2D patterns in an image and matching them to a 3D database of textured facades by reasoning about the underlying canonical forms of each pattern. Multiple 2D-to-3D pattern correspondences enable robust recovery of camera orientation and location. We demonstrate the success of this method in a large urban environment.

42. Recognition by Association via Learning Per-exemplar Distances

Tomasz Malisiewicz, Alexei A. Efros

We pose the recognition problem as data association. In this setting, a novel object is explained solely in terms of a small set of exemplar objects to which it is visually similar. Inspired by the work of Frome et al., we learn separate distance functions for each exemplar; however, our distances are interpretable on an absolute scale and can be thresholded to detect the presence of an object. Our exemplars are represented as image regions and the learned distances capture the relative importance of shape, color, texture, and position features for that region. We use the distance functions to detect and segment objects in novel images by associating the bottom-up segments obtained from multiple image segmentations with the exemplar regions. We evaluate the detection and segmentation performance of our algorithm on real-world outdoor scenes from the LabelMe [15] dataset and also show some promising qualitative image parsing results.

43. Improved Building Detection by Gaussian Processes Classification via Feature Space Rescale and Spectral Kernel Selection

Hang Zhou, David Suter

We use spectral analysis to facilitate Gaussian processes (GP) classification. Our solution provides two improvements: scaling of the data to achieve a more isotropic nature, as well as a method to choose the kernel to match certain data characteristics. Given the dataset, from the Fourier transform of the training data we compare the frequency domain features of each dimension to estimate a rescaling (towards making the data isotropic). Also, the spectrum of the training data is compared with several candidate kernel spectrums. From this comparison the best matching kernel is chosen. In these ways, the training data matches better the GP classification kernel function (and hence the underlying assumed correlation characteristics), resulting in a better GP classification result. Test results on both non image and image data show the efficiency and effectiveness of our approach.

44. A Scalable Graph-Cut Algorithm for N-D Grids

Andrew Delong, Yuri Boykov

Global optimization via s-t graph cuts is widely used in computer vision and graphics. To obtain high-resolution output, graph cut methods must construct massive N-D grid-graphs containing billions of vertices. We show that when these graphs do not fit into physical memory, current max-flow/min-cut algorithms--the workhorse of graph cut methods--are totally impractical. Others have resorted to banded or hierarchical approximation methods that get trapped in local minima, which loses the main benefit of global optimisation.

We enhance the push-relabel algorithm for maximum flow [14] with two practical contributions. First, true global minima can now be computed on immense grid-like graphs too large for physical memory. These graphs are ubiquitous in computer vision, medical imaging and graphics. Second, for commodity multi-core platforms our algorithm attains near-linear speedup with respect to number of processors. To achieve these goals, we generalised the standard relabeling operations associated with push-relabel.

45. Image Partial Blur Detection and Classification

Renting Liu, Zhaorong Li, Jiaya Jia

In this paper, we propose a partially-blurred-image classification and analysis framework for automatically detecting images containing blurred regions and recognizing the blur types for those regions without needing to perform blur kernel estimation and image deblurring. We develop several blur features modeled by image color, gradient, and spectrum information, and use feature parameter training to robustly classify blurred images. Our blur detection is based on image patches, making region-wise training and classification in one image efficient. Extensive experiments show that our method works satisfactorily on challenging image data, which establishes a technical foundation for solving several computer vision problems, such as motion analysis and image restoration, using the blur information.

46. Subspace segmentation with outliers: a Grassmannian approach to the maximum consensus subspace

Nuno Pinho da Silva, João Paulo Costeira

Segmenting arbitrary unions of linear subspaces is an important tool for computer vision tasks such as motion and image segmentation, SfM or object recognition. We segment subspaces by searching for the orthogonal complement of the subspace supported by the majority of the observations, i.e., the maximum consensus subspace. It is formulated as a grassmannian optimization problem: a smooth, constrained but nonconvex program is immersed into the Grassmann manifold, resulting in a low dimensional and unconstrained program solved with an efficient optimization algorithm. Nonconvexity implies that global optimality depends on the initialization. However, by finding the maximum consensus subspace, outlier rejection becomes an inherent property of the method. Besides robustness, it does not rely on prior global detection procedures (e.g., rank of data matrices), which is the case of most current works. We test our algorithm in both synthetic and real data, where no outlier was ever classified as inlier.

47. Robust Estimation of Gaussian Mixtures from Noisy Input Data

Shaobo Hou, Aphrodite Galata

We propose a variational bayes approach to the problem of robust estimation of gaussian mixtures from noisy input data. The proposed algorithm explicitly takes into account the uncertainty associated with each data point, makes no assumptions about the structure of the covariance matrices and is able to automatically determine the number of the gaussian mixture components. Through the use of both synthetic and real world data examples, we show that by incorporating uncertainty information into the clustering algorithm, we get better results at recovering the true distribution of the training data compared to other variational bayesian clustering algorithms.

48. Progressive Search Space Reduction for Human Pose Estimation

Vittorio Ferrari, Manuel J. Marín-Jiménez, Andrew Zisserman

The objective of this paper is to estimate 2D human pose as a spatial configuration of body parts in TV and movie video shots. Such video material is uncontrolled and extremely challenging.

We propose an approach that progressively reduces the search space for body parts, to greatly improve the chances that pose estimation will succeed. This involves two contributions: (i) a generic detector using a weak model of pose to substantially reduce the full pose search space; and (ii) employing `grabcut' initialized on detected regions proposed by the weak model, to further prune the search space. Moreover, we also propose (iii) an integrated spatiotemporal model covering multiple frames to refine pose estimates from individual frames, with inference using belief propagation.

The method is fully automatic and self-initializing, and explains the spatio-temporal volume covered by a person moving in a shot, by soft-labeling every pixel as belonging to a particular body part or to the background. We demonstrate upper-body pose estimation by an extensive evaluation over 70000 frames from four episodes of the TV series *Buffy the Vampire Slayer*, and present an application to full-body action recognition on the Weizmann dataset.

49. Branch-and-bound hypothesis selection for two-view multiple structure and motion segmentation

Ninad Thakoor, Jean Gao

An efficient and robust framework for two-view multiple structure and motion segmentation is proposed. To handle this otherwise recursive problem, hypotheses for the models are generated by local sampling. Once these hypotheses are available, a model selection problem is formulated which takes into account the hypotheses likelihoods and model complexity. An explicit model for outliers is also added for robust model selection. The model selection criterion is optimized through branch-and-bound technique of combinatorial optimization which guaranties optimality over current set of hypotheses by efficient search of solution space.

50. Graph Cut with Ordering Constraints on Labels and its Applications

Xiaoqing Liu, Olga Veksler, Jagath Samarabandu

In the last decade, graph-cut optimization has been popular for a variety of pixel labeling problems. Typically graph-cut methods are used to incorporate a smoothness prior on a labeling. Recently several methods incorporated ordering constraints on labels for the application of object segmentation. An example of an ordering constraint is prohibiting a pixel with a "car wheel" label to be above a pixel with a "car roof" label. We observe that the commonly used graph-cut based α- expansion is more likely to get stuck in a local minimum when ordering constraints are used. For certain models with ordering constraints, we develop new graph-cut moves which we call order-preserving moves. Order-preserving moves act on all labels, unlike α -expansion. Although the global minimum is still not guaranteed, optimization with order-preserving moves performs significantly better than α -expansion. We evaluate order-preserving moves for the geometric class scene labeling (introduced by Hoiem et al.) where the goal is to assign each pixel a label such as "sky", "ground", etc., so ordering constraints arise naturally. In addition, we use order-preserving moves for certain simple shape priors in graph-cut segmentation, which is a novel contribution in itself.

51. Superpixel Lattices

Alastair Moore, Simon Prince, Jonathan Warrell, Umar Mohammed, Graham Jones

Unsupervised over-segmentation of an image into superpixels is a common preprocessing step for image parsing algorithms. Ideally, every pixel within each superpixel region will belong to the same real-world object. Existing algorithms generate superpixels that forfeit many useful properties of the regular topology of the original pixels: for example, the n^{th} superpixel has no consistent position or relationship with its neighbors. We propose a novel algorithm that produces superpixels that are forced to conform to a grid (a regular superpixel lattice). Despite this added topological constraint, our algorithm is comparable in terms of speed and accuracy to alternative segmentation approaches. To demonstrate this, we use evaluation metrics based on (i) image reconstruction (ii) comparison to human-segmented images and (iii) stability of segmentation over subsequent frames of video sequences.

52. Object recognition and segmentation by non-rigid quasi-dense matching

Juho Kannala, Esa Rahtu, Sami S. Brandt, Janne Heikkilä

In this paper, we present a non-rigid quasi-dense matching method and its application to object recognition and segmentation. The matching method is based on the match propagation algorithm which is here extended by using local image gradients for adapting the propagation to smooth non-rigid deformations of the imaged surfaces. The adaptation is based entirely on the local properties of the images and the method can be hence used in non-rigid image registration where global geometric constraints are not available. Our approach for object recognition and segmentation is directly built on the quasi-dense matching. The quasi-dense pixel matches between the model and test images are grouped into geometrically consistent groups using a method which utilizes the local affine transformation estimates obtained during the propagation. The number and quality of geometrically consistent matches is used as a recognition criterion and the location of the matching pixels directly provides the segmentation. The experiments demonstrate that our approach is able to deal with extensive background clutter, partial occlusion, large scale and viewpoint changes, and notable geometric deformations.

53. Sequential Sparsification for Change Detection

Necmiye Ozay, Mario Sznaier, Octavia I. Camps

This paper presents a general method for segmenting a vector valued sequence into an unknown number of subsequences where all data points from a subsequence can be represented with the same affine parametric model. The idea is to cluster the data into the minimum number of such subsequences which, as we show, can be cast as a sparse signal recovery problem by exploiting the temporal correlation between consecutive data points. We try to maximize the sparsity (i.e., the number of zero elements) of the first order differences of the sequence of parameter vectors. Each non-zero element in the first order difference sequence corresponds to a change. A weighted l_1 norm based convex approximation is adopted to solve the change detection problem. We apply the proposed method to video segmentation and temporal segmentation of dynamic textures.

54. Efficient Sequential Correspondence Selection by Cosegmentation

Jan Čech, Jiří Matas, Michal Perd'och

In many retrieval, object recognition and wide baseline stereo methods, correspondences of interest points are established possibly sublinearly by matching a compact descriptor such as SIFT. We show that a subsequent cosegmentation process coupled with a quasi-optimal sequential decision process leads to a correspondence verification procedure that has (i) high precision (is highly discriminative) (ii) good recall and (iii) is fast. The sequential decision on the correctness of a correspondence is based on trivial attributes of a modified dense stereo matching algorithm. The attributes are projected on a prominent discriminative direction by SVM. Wald's sequential probability ratio test is performed for SVM projection computed on progressively larger co-segmented regions. Experimentally we show that the process significantly outperforms the standard correspondence selection process based on SIFT distance ratios on challenging matching problems.

55. A Multicompartment Segmentation Framework With Homeomorphic Level Sets

Xian Fan, Pierre-Louis Bazin, Jerry Prince

The simultaneous segmentation of multiple objects is an important problem in many imaging and computer vision applications. Various extensions of level set segmentation techniques to multiple objects have been proposed; however, no one method maintains object relationships, preserves topology, is computationally efficient, and provides an object-dependent internal and external force capability. In this paper, a framework for segmenting multiple objects that permits different forces to be applied to different boundaries while maintaining object topology and relationships is presented. Because of this framework, the segmentation of multiple objects each with multiple compartments is supported, and no overlaps or vacuums are generated. The computational complexity of this approach is independent of the number of objects to segment, thereby permitting the simultaneous segmentation of a large number of components. The properties of this approach and comparisons to existing methods are shown using a variety of images, both synthetic and real.

56. Image Segmentation with a Parametric Deformable Model Using Shape and Appearance Priors

Ayman El-Baz, Georgy Gimel'farb

We propose a novel parametric deformable model controlled by shape and visual appearance priors learned from a training subset of co-aligned images of goal objects. The shape prior is derived from a linear combination of vectors of distances between the training boundaries and their common centroid. The appearance prior considers gray levels within each training boundary as a sample of a Markov-Gibbs random field with pairwise interaction. Spatially homogeneous interaction geometry and Gibbs potentials are analytically estimated from the training data. To accurately separate a goal object from an arbitrary background, empirical marginal gray level distributions inside and outside of the boundary are modeled with adaptive linear combinations of discrete Gaussians (LCDG). The evolution of the parametric deformable model is based on solving an Eikonal partial differential equation with a new speed function which combines the prior shape, prior appearance, and current appearance models. Due to the analytical shape and appearance priors and a simple Expectation-Maximization procedure for getting the object and background LCDG, our segmentation is considerably faster than most of the known geometric and parametric models. Experiments with various goal images confirm the robustness, accuracy, and speed of our approach.

57. Towards Unsupervised Whole-Object Segmentation: Combining Automated Matting with Boundary Detection

Andrew N. Stein, Thomas S. Stepleton, Martial Hebert

We propose a novel step toward the unsupervised segmentation of whole objects by combining “hints” of partial scene segmentation offered by multiple soft, binary mattes. These mattes are implied by a set of hypothesized object boundary fragments in the scene. Rather than trying to find or define a single “best” segmentation, we generate multiple segmentations of an image. This reflects contemporary methods for unsupervised object discovery from groups of images, and it allows us to define intuitive evaluation metrics for our sets of segmentations based on the accurate and parsimonious delineation of scene objects. Our proposed approach builds on recent advances in spectral clustering, image matting, and boundary detection. It is demonstrated qualitatively and quantitatively on a dataset of scenes and is suitable for current work in unsupervised object discovery without top-down knowledge.

58. Sparsity, Redundancy and Optimal Image Support towards Knowledge-based Segmentation

Salma Essafi, Georg Langs, Nikos Paragios

In this paper, we propose a novel approach to model shape variations. It encodes sparsity, exploits geometric redundancy, and accounts for the different degrees of local variation and image support. In this context we consider a control-point based shape representation. Their sparse distribution is derived based on a shape model metric learned from the training data, and the ambiguity of local appearance with regard to segmentation changes. The resulting sparse model of the object improves reconstruction and search behavior, in particular for data that exhibit a heterogeneous distribution of image information and shape complexity. Furthermore, it goes beyond conventional imagebased segmentation approaches since it is able to identify reliable image structures which are then encoded within the model and used to determine the optimal segmentation map. We report promising experimental results comparing our approach with standard models on MRI data of calf muscles - an application where traditional image-based methods fail - and CT data of the left heart ventricle.

59. Modeling the structure of multivariate manifolds: Shape Maps

Georg Langs, Nikos Paragios

We propose a shape population metric that reflects the interdependencies between points observed in a set of examples. It provides a notion of topology for shape and appearance models that represents the behavior of individual observations in a metric space, in which distances between points correspond to their joint modeling properties. A Markov chain is learnt using the description lengths of models that describe sub sets of the entire data. The according diffusion map or shape map provides for the metric that reflects the behavior of the training population. With this metric functional clustering, deformation- or motion segmentation, sparse sampling and the treatment of outliers can be dealt with in a unified and transparent manner. We report experimental results on synthetic and real world data and compare the framework with existing specialized approaches.

60. Conditional Density Learning via Regression with Application to Deformable Shape Segmentation

Jingdan Zhang, Shaohua Zhou, Dorin Comaniciu, Leonard McMillan

Many vision problems can be cast as optimizing the conditional probability density function $p(C|I)$ where I is an image and C is a vector of model parameters describing the image. Ideally, the density function $p(C|I)$ would be smooth and unimodal allowing local optimization techniques, such as gradient descent or simplex, to converge to an optimal solution quickly, while preserving significant nonlinearities of the model. We propose to learn a conditional probability density satisfying these desired properties for the given training data set. To do this, we formulate a novel regression problem that finds a function approximating the target density. Learning the regressor is challenging due to the high dimensionality of model parameters, C, and the complexity of relating the image and the model. Our approach makes two contributions. First, we take a multi-level refinement approach by learning a series of density functions, each of which guides the solution of optimization algorithms increasingly converging to the correct solution. Second, we propose a new data sampling algorithm that takes into account the gradient information of the target function. We have applied this learning approach to deformable shape segmentation and have achieved better accuracy than the previous methods.

61. Multi-Image Graph Cut Clothing Segmentation for Recognizing People

Andrew C. Gallagher, Tsuhan Chen

Reseachers have verified that clothing provides information about the identity of the individual. To extract features from the clothing, the clothing region first must be localized or segmented in the image. At the same time, given multiple images of the same person wearing the same clothing, we expect to improve the effectiveness of clothing segmentation. Therefore, the identity recognition and clothing segmentation problems are inter-twined; a good solution for one aides in the solution for the other.

We build on this idea by analyzing the mutual information between pixel locations near the face and the identity of the person to learn a global clothing mask. We segment the clothing region in each image using graph cuts based on a clothing model learned from one or multiple images believed to be the same person wearing the same clothing. We use facial features and clothing features to recognize individuals in other images. The results show that clothing segmentation provides a significant improvement in recognition accuracy for large image collections, and useful clothing masks are simultaneously produced.

A further significant contribution is that we introduce a publicly available consumer image collection where each individual is identified. We hope this dataset allows the vision community to more easily compare results for tasks related to recognizing people in consumer image collections.

62. Smoothing-based Optimization

Leordeanu Marius, Hebert Martial

We propose an efficient method for complex optimization problems that often arise in computer vision. While our method is general and could be applied to various tasks, it was mainly inspired from problems in computer vision, and it borrows ideas from scale space theory. One of the main motivations for our approach is that searching for the global maximum through the scale space of a function is equivalent to looking for the maximum of the original function, with the advantage of having to handle fewer local optima. Our method works with any non-negative, possibly non-smooth function, and requires only the ability of evaluating the function at any specific point. The algorithm is based on a growth transformation, which is guaranteed to increase the value of the scale space function at every step, unlike gradient methods. To demonstrate its effectiveness we present its performance on a few computer vision applications, and show that in our experiments it is more effective than some well established methods such as MCMC, Simulated Annealing and the more local Nelder-Mead optimization method.

63. The Scale of a Texture and its Application to Segmentation

Byung-Woo Hong, Stefano Soatto, Kangyu Ni, Tony Chan

This paper examines the issue of scale in modeling texture for the purpose of segmentation. We propose a scale descriptor for texture and an energy minimization model to find the scale of a given texture at each location. For each pixel, we use the intensity distribution in a local patch around that pixel to determine the smallest size of the domain that can be used to generate neighboring patches. The energy functional we propose to minimize is comprised of three terms: The first is the dissimilarity measure using the Wasserstein distance or Kullback-Leibler divergence between neighboring patch distributions; the second maximizes the entropy of the local patch, and the third penalizes larger size at equal fidelity. Our experiments show the proposed scale model successfully captures the intrinsic scale of texture at each location. We also apply our scale descriptor for improving texture segmentation based on histogram matching [15].

64. A probabilistic segmentation method for the identification of luminal borders in intravascular ultrasound images

Gerardo Mendizabal-Ruiz, Mariano Rivera, Ioannis A. Kakadiaris

Intravascular ultrasound (IVUS) is a catheter-based medical imaging technique that produces cross-sectional images of blood vessels and is particularly useful for studying atherosclerosis. In this paper, we present a probabilistic approach for the semi-automatic identification of the luminal border on IVUS images. Specifically, we parameterize the lumen contour using a mixture of Gaussian that is deformed by the minimization of a cost function formulated using a probabilistic approach. For the optimization of the cost function, we introduce a novel method that linearly combines the descent directions of the steepest descent and BFGS optimization methods within a trust region that improves convergence. Results of our proposed method on 20 MHz IVUS images are presented and discussed in order to demonstrate the effectiveness of our approach.

65. Interactive Image Segmentation Via Minimization of Quadratic Energies on Directed Graphs

Dheeraj Singaraju, Leo Grady, René Vidal

We propose a scheme to introduce directionality in the Random Walker algorithm for image segmentation. In particular, we extend the optimization framework of this algorithm to combinatorial graphs with directed edges. Our scheme is interactive and requires the user to label a few pixels that are representative of a foreground object and of the background. These labeled pixels are used to learn intensity models for the object and the background, which allow us to automatically set the weights of the directed edges. These weights are chosen so that they bias the direction of the object boundary gradients to flow from regions that agree well with the learned object intensity model to regions that do not agree well. We use these weights to define an energy function that associates asymmetric quadratic penalties with the edges in the graph. We show that this energy function is convex, hence it has a unique minimizer. We propose a provably convergent iterative algorithm for minimizing this energy function. We also describe the construction of an equivalent electrical network with diodes and resistors that solves the same segmentation problem as our framework. Finally, our experiments on a database of 69 images show that the use of directional information does improve the segmenting power of the Random Walker algorithm.

66. Joint Tracking of Features and Edges

Stan Birchfield, Shrinivas Pundlik

Sparse features have traditionally been tracked from frame to frame independently of one another. We propose a framework in which features are tracked jointly. Combining ideas from Lucas-Kanade and Horn-Schunck, the estimated motion of a feature is influenced by the estimated motion of neighboring features. The approach also handles the problem of tracking edges in a unified way by estimating motion perpendicular to the edge, using the motion of neighboring features to resolve the aperture problem. Results are shown on several image sequences to demonstrate the improved results obtained by the approach.

67. Fast Approximate Random Walker Segmentation Using Eigenvector Precomputation

Leo Grady, Sinop Ali

Interactive segmentation is often performed on images that have been stored on disk (e.g., a medical image server) for some time prior to user interaction. We propose to use this time to perform an offline precomputation of the segmentation prior to user interaction that significantly decreases the amount of user time necessary to produce a segmentation. Knowing how to effectively precompute the segmentation prior to user interaction is difficult, since a user may choose to guide the segmentation algorithm to segment any object (or multiple objects) in the image. Consequently, precomputation performed prior to user interaction must be performed without any knowledge of the user interaction. Specifically, we show that one may precompute several eigenvectors of the weighted Laplacian matrix of a graph and use this information to produce a linear-time approximation of the Random Walker segmentation algorithm, even without knowing where the foreground/background seeds will be placed. Finally, we also show that this procedure may be interpreted as a seeded (interactive) Normalized Cuts algorithm.

68. Detection and Matching of Rectilinear Structures

Branislav Mičušík, Horst Wildenauer, Jana Košecká

Indoor and outdoor urban environments posses many regularities which can be efficiently exploited and used for general image parsing tasks. We present a novel approach for detecting rectilinear structures and demonstrate their use for wide baseline stereo matching, planar 3D reconstruction, and computation of geometric context. Assuming a presence of dominant orthogonal vanishing directions, we proceed by formulating the detection of the rectilinear structures as a labeling problem on detected line segments. The line segment labels, respecting the proposed grammar rules, are established as the MAP assignment of the corresponding MRF. The proposed framework allows to detect both full as well as partial rectangles, rectangle-in-rectangle structures, and rectangles sharing edges. The use of detected rectangles is demonstrated in the context of difficult wide baseline matching tasks in the presence of repetitive structures and large appearance changes.

69. Discriminative Modeling by Boosting on Multilevel Aggregates

Jason J. Corso

This paper presents a new approach to discriminative modeling for classification and labeling. Our method, called Boosting on Multilevel Aggregates (BMA), adds a new class of hierarchical, adaptive features into boosting-based discriminative models. Each pixel is linked with a set of aggregate regions in a multilevel coarsening of the image. The coarsening is adaptive, rapid and stable. The multilevel aggregates present additional information rich features on which to boost, such as shape properties, neighborhood context, hierarchical characteristics, and photometric statistics. We implement and test our approach on three two-class problems: classifying documents in office scenes, buildings and horses in natural images. In all three cases, the majority, about 75%, of features selected during boosting are our proposed BMA features rather than patch-based features. This large percentage demonstrates the discriminative power of the multilevel aggregate features over conventional patch-based features. Our quantitative performance measures show the proposed approach gives superior results to the state-of-the-art in all three applications.

70. Graph-Shifts: Natural Image Labeling by Dynamic Hierarchical Computing

Jason J. Corso, Alan Yuille, Zhuowen Tu

In this paper, we present a new approach for image labeling based on the recently introduced graph-shifts algorithm. Graph-shifts is an energy minimization algorithm that does labeling by dynamically manipulating, or shifting, the parent-child relationships in a hierarchical decomposition of the image. Each shift optimally reduces the energy by indirectly causing a change to the labeling; graph-shifts is able to rapidly compute and select this optimal shift at every iteration. There are no constraints on the terms of the (pairwise) energy function. The algorithm was originally presented in the context of medical image labeling using conditional random field models. In this paper, we consider the algorithm in the context of both low- and high-level natural image labeling. We show that for examples in both classes of problems, graph-shifts does labeling both accurately and rapidly. For low-level vision, we explore image restoration, and for high-level vision, we make use of a hybrid discriminative-generative model to segment and label images into semantically meaningful regions (e.g., trees, buildings, etc.). For both problems, we obtain comparable or superior results to the state-of-the-art computed in just a few seconds per image.

71. A Bi-illuminant Dichromatic Reflection Model for Understanding Images

Bruce A. Maxwell, Richard M. Friedhoff, Casey A. Smith

This paper presents a new model for understanding the appearance of objects that exhibit both body and surface reflection under realistic illumination. Specifically, the model represents the appearance of surfaces that interact with a dominant illuminant and a non-negligible ambient illuminant that may have different spectral power distributions. Real illumination environments usually have an ambient illuminant, and the current dynamic range of consumer cameras is sufficient to capture significant information in shadows. The bi-illuminant dichromatic reflection model explains numerous empirical findings in the literature and has implications for commonly used chromaticity spaces that claim to be illumination invariant but are not in many natural situations. One outcome of the model is the first 2-D chromaticity space for an RGB image that is robust to illumination change given dominant and ambient illuminants with different spectral power distributions.

72. Constrained Image Segmentation from Hierarchical Boundaries

Pablo Arbeláez, Laurent Cohen

In this paper, we address the problem of constrained segmentation of natural images, in which a human user places one seed point inside each object of interest in the image and the task is to determine the object boundaries. For this purpose, we study the connection between seed-based and hierarchical segmentation. We consider an Ultrametric Contour Map (UCM), the representation of a hierarchy of segmentations as a real-valued boundary image. Starting from a set of seed points, we propose an algorithm for constructing Voronoi tessellations with respect to a distance defined by the UCM. As a result, the main contribution of the paper is a method that allows exploiting the information of any hierarchical scheme for constrained segmentation. Our algorithm is parameter-free, computationally efficient and robust. We prove the interest of the approach proposed by evaluating quantitatively the results with respect to ground-truth data.

73. Image De-fencing

Yanxi Liu, Tamara Belkina, James H. Hays, Roberto Lublinerman

We introduce a novel image segmentation algorithm that uses translational symmetry as the primary foreground/background separation cue. We investigate the process of identifying and analyzing image regions that present approximate translational symmetry for the purpose of image foreground/background separation. In conjunction with texture-based inpainting, understanding the different see-through layers allows us to perform powerful image manipulations such as recovering a mesh-occluded background (as much as 53% occluded area) to achieve the effect of image and photo de-fencing. Our algorithm consists of three distinct phases-- (1) automatically finding the skeleton structure of a potential frontal layer (fence) in the form of a deformed lattice, (2) separating foreground/background layers using appearance regularity, and (3) occluded foreground inpainting to reveal a complete, non-occluded image. Each of these three tasks presents its own special computational challenges that are not encountered in previous, general image de-layering or texture inpainting applications.

74. Extracting Smooth and Transparent Layers from a Single Image

Sai-Kit Yeung, Tai-Pang Wu, Chi-Keung Tang

Layer decomposition from a single image is an under-constrained problem, because there are more unknowns than equations. This paper studies a slightly easier but very useful alternative where only the background layer has substantial image gradients and structures. We propose to solve this useful alternative by an expectation-maximization (EM) algorithm that employs the hidden markov model (HMM), which maintains spatial coherency of smooth and overlapping layers, and helps to preserve image details of the textured background layer. We demonstrate that, using a small amount of user input, various seemingly unrelated problems in computational photography can be effectively addressed by solving this alternative using our EM-HMM algorithm.

3:45pm – 5:30pm Oral Session O1P-1: Motion Analysis for Structure, Shape and Pose (Arteaga)

1. Dense 3D Motion Capture from Synchronized Video Streams

Yasutaka Furukawa, Jean Ponce

This paper proposes a novel approach to nonrigid, markerless motion capture from synchronized video streams acquired by calibrated cameras. The instantaneous geometry of the observed scene is represented by a polyhedral mesh with fixed topology. The initial mesh is constructed in the first frame using the publicly available PMVS software for multi-view stereo [7]. Its deformation is captured by tracking its vertices over time, using two optimization processes at each frame: a local one using a rigid motion model in the neighborhood of each vertex, and a global one using a regularized nonrigid model for the whole mesh. Qualitative and quantitative experiments using seven real datasets show that our algorithm effectively handles complex nonrigid motions and severe occlusions.

2. Recovering Consistent Video Depth Maps via Bundle Optimization

Guofeng Zhang, Jiaya Jia, Tien-Tsin Wong, Hujun Bao

This paper presents a novel method for reconstructing high-quality video depth maps. A bundle optimization model is proposed to address the key issues, including image noise and occlusions, in stereo reconstruction. Our method not only uses the color constancy constraint, but also explicitly incorporates the geometric coherence constraint associating multiple frames in a video, thus can naturally maintain the temporal coherence of the recovered video depths without introducing over-smoothing artifact. To make the inference problem tractable, we introduce an iterative optimization scheme by first initializing disparity maps using segmentation prior and then refining the disparities by means of bundle optimization. Unlike previous work estimating complex visibility parameters, our approach implicitly models the probabilistic visibility in a statistical way. The effectiveness of our automatic method is demonstrated using challenging video examples.

3. Directions of Egomotion from Antipodal Points

John Lim, Nick Barnes

We present a novel geometrical constraint on the egomotion of a single, moving camera. Using a camera with a large field-of-view (FOV), the optical flow measured at a single pair of antipodal points on the image sphere constrains the set of all possible camera motion directions to a subset region. By considering the flow at many such antipodal point pairs, it is shown that the intersection of all subset regions arising from each pair yields an estimate on the directions of motion. These antipodal point constraints rely on the geometrical properties of using a spherical representation of the image as well as the larger information content available from a large FOV. An algorithm using these constraints was implemented and tested on both simulated and real images. Results show comparable performance to the state of the art in the presence of noise and outliers whilst processing in constant time.

4. 3D Pose Refinement from Reflections

Pascal Lagger, Mathieu Salzmann, Vincent Lepetit, Pascal Fua

We demonstrate how to exploit reflections for accurate registration of shiny objects: The lighting environment can be retrieved from the reflections under a distant illumination assumption. Since it remains unchanged when the camera or the object of interest moves, this provides powerful additional constraints that can be incorporated into standard pose estimation algorithms.

The key idea and main contribution of the paper is therefore to show that the registration should also be performed in the lighting environment space, instead of in the image space only. This lets us recover very accurate pose estimates because the specularities are very sensitive to pose changes. An interesting side result is an accurate estimate of the lighting environment.

Furthermore, since the mapping from lighting environment to specularities has no analytical expression for objects represented as 3D meshes, and is not 1-to-1, registering lighting environments is far from trivial. However we propose a general and effective solution. Our approach is demonstrated on both synthetic and real images.

5. Local Deformation Models for Monocular 3D Shape Recovery

Mathieu Salzmann, Raquel Urtasun, Pascal Fua

Without a deformation model, monocular 3D shape recovery of deformable surfaces is severly under-constrained. Even when the image information is rich enough, prior knowledge of the feasible deformations is required to overcome the ambiguities. This is further accentuated when such information is poor, which is a key issue that has not yet been addressed.

In this paper, we propose an approach to learning shape priors to solve this problem. By contrast with typical statistical learning methods that build models for specific object shapes, we learn local deformation models, and combine them to reconstruct surfaces of arbitrary global shapes. Not only does this improve the generality of our deformation models, but it also facilitates learning since the space of local deformations is much smaller than that of global ones.

While using a texture-based approach, we show that our models are effective to reconstruct from single videos poorly-textured surfaces of arbitrary shape, made of materials as different as cardboard, that deforms smoothly, and much lighter tissue paper whose deformations may be far more complex.

3:45pm – 5:30pm Oral Session O1P-2 : Object Detection, Categorization and Recognition (Cook)

1. Probabilistic Graph and Hypergraph Matching

Ron Zass, Amnon Shashua

We consider the problem of finding a matching between two sets of features, given complex relations among them, going beyond pairwise. Each feature set is modeled by a hypergraph where the complex relations are represented by hyper-edges. A match between the feature sets is then modeled as a hypergraph matching problem. We derive the hyper-graph matching problem in a probabilistic setting represented by a convex optimization. First, we formalize a soft matching criterion that emerges from a probabilistic interpretation of the problem input and output, as opposed to previous methods that treat soft matching as a mere relaxation of the hard matching problem. Second, the model induces an algebraic relation between the hyper-edge weight matrix and the desired vertex-to-vertex probabilistic matching. Third, the model explains some of the graph matching normalization proposed in the past on a heuristic basis such as doubly stochastic normalizations of the edge weights. A key benefit of the model is that the global optimum of the matching criteria can be found via an iterative successive projection algorithm. The algorithm reduces to the well known Sinkhorn [15] row/column matrix normalization procedure in the special case when the two graphs have the same number of vertices and a complete matching is desired. Another benefit of our model is the straightforward scalability from graphs to hyper-graphs.

2. 3D Model Matching with Viewpoint-Invariant Patches (VIPs)

Changchang Wu, Brian Clipp, Xiaowei Li, Jan-Michael Frahm, Marc Pollefeys

The robust alignment of images and scenes seen from widely different viewpoints is an important challenge for camera and scene reconstruction. This paper introduces a novel class of viewpoint independent local features for robust registration and novel algorithms to use the rich information of the new features for 3D scene alignment and large scale scene reconstruction. The key point of our approach consists of leveraging local shape information for the extraction of an invariant feature descriptor. The advantages of the novel viewpoint invariant patch (VIP) are: that the novel features are invariant to 3D camera motion and that a single VIP correspondence uniquely defines the 3D similarity transformation between two scenes. In the paper we demonstrate how to use the properties of the VIPs in an efficient matching scheme for 3D scene alignment. The algorithm is based on a hierarchical matching method which tests the components of the similarity transformation sequentially to allow efficient matching and 3D scene alignment. We evaluate the novel features on real data with known ground truth information and show that the features can be used to reconstruct large scale urban scenes.

3. Unsupervised Modeling of Object Categories Using Link Analysis Techniques

Gunhee Kim, Christos Faloutsos, Martial Hebert

We propose an approach for learning visual models of object categories in an unsupervised manner in which we first build a large-scale complex network which captures the interactions of all unit visual features across the entire training set and we infer information, such as which features are in which categories, directly from the graph by using link analysis techniques. The link analysis techniques are based on well-established graph mining techniques used in diverse applications such as WWW, bioinformatics, and social networks. The techniques operate directly on the patterns of connections between features in the graph rather than on statistical properties, e.g., from clustering in feature space. We argue that the resulting techniques are simpler, and we show that they perform similarly or better compared to state of the art techniques on common data sets. We also show results on more challenging data sets than those that have been used in prior work on unsupervised modeling.

4. Semantic Texton Forests for Image Categorization and Segmentation

Jamie Shotton, Matthew Johnson, Roberto Cipolla

We propose semantic texton forests, efficient and powerful new low-level features. These are ensembles of decision trees that act directly on image pixels, and therefore do not need the expensive computation of filter-bank responses or local descriptors. They are extremely fast to both train and test, especially compared with k-means clustering and nearest-neighbor assignment of feature descriptors. The nodes in the trees provide (i) an implicit hierarchical clustering into semantic textons, and (ii) an explicit local classification estimate. Our second contribution, the bag of semantic textons, combines a histogram of semantic textons over an image region with a region prior category distribution. The bag of semantic textons is computed over the whole image for categorization, and over local rectangular regions for segmentation. Including both histogram and region prior allows our segmentation algorithm to exploit both textural and semantic context. Our third contribution is an image-level prior for segmentation that emphasizes those categories that the automatic categorization believes to be present. We evaluate on two datasets including the very challenging VOC 2007 segmentation dataset. Our results significantly advance the state-of-the-art in segmentation accuracy, and furthermore, our use of efficient decision forests gives at least a five-fold increase in execution speed.

5. Unifying Discriminative Visual Codebook Generation with Classifier Training for Object Category Recognition

Liu Yang, Rong Jin, Rahul Sukthankar, Frédéric Jurie

The idea of representing images using a bag of visual words is currently popular in object category recognition. Since this representation is typically constructed using unsupervised clustering, the resulting visual words may not capture the desired information. Recent work has explored the construction of discriminative visual codebooks that explicitly consider object category information. However, since the codebook generation process is still disconnected from that of classifier training, the set of resulting visual words, while individually discriminative, may not be those best suited for the classifier. This paper proposes a novel optimization framework that unifies codebook generation with classifier training. In our approach, each image feature is encoded by a sequence of "visual bits" optimized for each category. An image, which can contain objects from multiple categories, is represented using aggregates of visual bits for each category. Classifiers associated with different categories determine how well a given image corresponds to each category. Based on the performance of these classifiers on the training data, we augment the visual words by generating additional bits. The classifiers are then updated to incorporate the new representation. These two phases are repeated until the desired performance is achieved. Experiments compare our approach to standard clustering-based methods and with state-of-the-art discriminative visual codebook generation. The significant improvements over previous techniques clearly demonstrate the value of unifying representation and classification into a single optimization framework.

8:30am – 10:30am Poster Session P2A-1: Motion and Tracking (Summit Hall)

1. An Adaptive Learning method for Target Tracking across Multiple Cameras

Kuan-Wen Chen, Chih-Chuan Lai, Yi-Ping Hung, Chu-Song Chen

This paper proposes an adaptive learning method for tracking targets across multiple cameras with disjoint views. Two visual cues are usually employed for tracking targets across cameras: spatio-temporal cue and appearance cue. To learn the relationships among cameras, traditional methods used batch-learning procedures or hand-labeled correspondence, which can work well only within a short period of time. In this paper, we propose an unsupervised method which learns both spatio-temporal relationships and appearance relationships adaptively and can be applied to long-term monitoring. Our method performs target tracking across multiple cameras while also considering the environment changes, such as sudden lighting changes. Also, we improve the estimation of spatio-temporal relationships by using the prior knowledge of camera network topology.

2. General Constraints for Batch Multiple-Target Tracking Applied to Large-Scale Videomicroscopy

Kevin Smith, Alan Carleton, Vincent Lepetit

While there is a large class of Multiple-Target Tracking (MTT) problems for which batch processing is possible and desirable, batch MTT remains relatively unexplored in comparison to sequential approaches. In this paper, we give a principled probabilistic formalization of batch MTT in which we introduce two new, very general constraints that considerably help us in reaching the correct solution. First, we exploit the correlation between the appearance of a target and its motion. Second, entrances and departures of targets are encouraged to occur at the boundaries of the scene. We show how to implement these constraints in a formal and efficient manner.

Our approach is applied to challenging 3-D biomedical imaging data where the number of targets is unknown and may vary, and numerous challenging tracking events occur. We demonstrate the ability of our model to simultaneously track the nuclei of over one hundred migrating neuron precursor cells in image stack series collected from a 2-photon microscope.

3. Image/Video Deblurring using a Hybrid Camera

Yu-Wing Tai, Hao Du, Michael Brown, Stephen Lin

We propose a novel approach to reduce spatially varying motion blur using a hybrid camera system that simultaneously captures high-resolution video at a low-frame rate together with low-resolution video at a high-frame rate. Our work is inspired by Ben-Ezra and Nayar who introduced the hybrid camera idea for correcting global motion blur for a single still image. We broaden the scope of the problem to address spatially varying blur as well as video imagery. We also reformulate the correction process to use more information available in the hybrid camera system, as well as iteratively refine spatially varying motion extracted from the low-resolution high-speed camera. We demonstrate that our approach achieves superior results over existing work and can be extended to deblurring of moving objects.

4. Efficient Mean Shift Belief Propagation for Vision Tracking

Minwoo Park, Yanxi Liu, Robert Collins

A mechanism for efficient mean-shift belief propagation (MSBP) is introduced. The novelty of our work is to use mean-shift to perform nonparametric mode-seeking on belief surfaces generated within the belief propagation framework. Belief Propagation (BP) is a powerful solution for performing inference in graphical models. However, there is a quadratic increase in the cost of computation with respect to the size of the hidden variable space. While the recently proposed nonparametric belief propagation (NBP) has better performance in terms of speed, even for continuous hidden variable spaces, computation is still slow due to the particle filter sampling process. Our MSBP method only needs to compute a local grid of samples of the belief surface during each iteration. This approach needs a significantly smaller number of samples than NBP, reducing computation time, yet it also yields more accurate and stable solutions. The efficiency and robustness of MSBP is compared against other variants of BP on applications in multi-target tracking and 2D articulated body tracking.

5. Visual Quasi-Periodicity

Erik Pogalin, Arnold W. M. Smeulders, Andrew H. C. Thean

Periodicity is at the core of the recognition of many actions. This paper takes the following steps to detect and measure periodicity. 1) We establish a conceptual framework of classifying periodicity in 10 essential cases, the most important of which are flashing (of a traffic light), pulsing (of an anemone), swinging (of wings), spinning (of a swimmer), turning (of a conductor), shuttling (of a brush), drifting (of an escalator) and thrusting (of a kangaroo). 2) We present an algorithm to detect all cases by the one and the same algorithm. It tracks the object independent of the object's appearance, then performs probabilistic PCA and spectral analysis followed by detection and frequency measurement. The method shows good performance with fixed parameters for examples of all above cases assembled from the Internet. 3) Application of the method, completely unaltered, to a random half hour of CNN news has led to an 80% score.

6. Learning Object Motion Patterns for Anomaly Detection and Improved Object Detection

Arslan Basharat, Alexei Gritai, Mubarak Shah

We present a novel framework for learning patterns of motion and sizes of objects in static camera surveillance. The proposed method provides a new higher-level layer to the traditional surveillance pipeline for anomalous event detection and scene model feedback. Pixel level probability density functions (pdfs) of appearance have been used for background modelling in the past, but modelling pixel level pdfs of object speed and size from the tracks is novel. Each pdf is modelled as a multivariate Gaussian Mixture Model (GMM) of the motion (destination location & transition time) and the size (width & height) parameters of the objects at that location. Output of the tracking module is used to perform unsupervised EM-based learning of every GMM. We have successfully used the proposed scene model to detect local as well as global anomalies in object tracks. We also show the use of this scene model to improve object detection through pixel-level parameter feedback of the minimum object size and background learning rate. Most object path modelling approaches first cluster the tracks into major paths in the scene, which can be a source of error. We avoid this by building local pdfs that capture a variety of tracks which are passing through them. Qualitative and quantitative analysis of actual surveillance videos proved the effectiveness of the proposed approach.

7. Context and Observation Driven Latent Variable Model for Human Pose Estimation

Abhinav Gupta, Trista Chen, Francine Chen, Don Kimber, Larry Davis

Current approaches to pose estimation and tracking can be classified into two categories: generative and discriminative. While generative approaches can accurately determine human pose from image observations, they are computationally expensive due to search in the high dimensional human pose space. On the other hand, discriminative approaches do not generalize well, but are computationally efficient. We present a hybrid model that combines the strengths of the two in an integrated learning and inference framework. We extend the Gaussian process latent variable model (GPLVM) to include an embedding from observation space (the space of image features) to the latent space. GPLVM is a generative model, but the inclusion of this mapping provides a discriminative component, making the model observation driven. Observation Driven GPLVM (OD-GPLVM) not only provides a faster inference approach, but also more accurate estimates (compared to GPLVM) in cases where dynamics are not sufficient for the initialization of search in the latent space.

We also extend OD-GPLVM to learn and estimate poses from parameterized actions/gestures. Parameterized gestures are actions which exhibit large systematic variation in joint angle space for different instances due to difference in contextual variables. For example, the joint angles in a forehand tennis shot are function of the height of the ball (Figure 2). We learn these systematic variations as a function of the contextual variables. We then present an approach to use information from scene/objects to provide context for human pose estimation for such parameterized actions.

8. Sequential Particle Swarm Optimization for Visual Tracking

Xiaoqin Zhang, Weiming Hu, Steve Maybank, Xi Li, Mingliang Zhu

Visual tracking usually involves an optimization process for estimating the motion of an object from measured images in a video sequence. In this paper, a new evolutionary approach, PSO (particle swarm optimization), is adopted for visual tracking. Since the tracking process is a dynamic optimization problem which is simultaneously influenced by the object state and the time, we propose a sequential particle swarm optimization framework by incorporating the temporal continuity information into the traditional PSO algorithm. In addition, the parameters in PSO are changed adaptively according to the fitness values of particles and the predicted motion of the tracked object, leading to a favourable performance in tracking applications. Furthermore, we show theoretically that, in a Bayesian inference view, the sequential PSO framework is in essence a multi-layer importance sampling based particle filter. Experimental results demonstrate that, compared with the state-of-the-art particle filter and its variation-the unscented particle filter, the proposed tracking algorithm is more robust and effective, especially when the object has an arbitrary motion or undergoes large appearance changes.

9. 3D Occlusion Recovery using Few Cameras

Mark Keck, James Davis

We present a practical framework for detecting and modeling 3D static occlusions for wide-baseline, multi-camera scenarios where the number of cameras is small. The framework consists of an iterative learning procedure where at each frame the occlusion model is used to solve the voxel occupancy problem, and this solution is then used to update the occlusion model. Along with this iterative procedure, there are two contributions of the proposed work: (1) a novel energy function (which can be minimized via graph cuts) specifically designed for use in this procedure, and (2) an application that incorporates our probabilistic occlusion model into a 3D tracking system. Both qualitative and quantitative results of the proposed algorithm and its incorporation with a 3D tracker are presented for support.

10. Online Learning of Patch Perspective Rectification for Efficient Object Detection

Stefan Hinterstoisser, Selim Benhimane, Nassir Navab, Pascal Fua, Vincent Lepetit

For a large class of applications, there is time to train the system. In this paper, we propose a learning-based approach to patch perspective rectification, and show that it is both faster and more reliable than state-of-the-art ad hoc affine region detection methods.

Our method performs in three steps. First, a classifier provides for every keypoint not only its identity, but also a first estimate of its transformation. This estimate allows carrying out, in the second step, an accurate perspective rectification using linear predictors. We show that both the classifier and the linear predictors can be trained online, which makes the approach convenient. The last step is a fast verification--made possible by the accurate perspective rectification--of the patch identity and its sub-pixel precision position estimation. We test our approach on real-time 3D object detection and tracking applications. We show that we can use the estimated perspective rectifications to determine the object pose and as a result, we need much fewer correspondences to obtain a precise pose estimation.

11. Sensor Planning for Automated and Persistent Object Tracking with Multiple Cameras

Yi Yao, Chung-hao Chen, Besma Abidi, David Page, Andreas Koschan, Mongi Abidi

Most existing camera placement algorithms focus on coverage and/or visibility analysis, which ensures that the object of interest is visible in the camera's field of view (FOV). However, visibility, a fundamental requirement of object tracking, is insufficient for persistent and automated tracking. In such applications, a continuous and consistently labeled trajectory of the same object should be maintained across different cameras' views. Therefore, a sufficient overlap between the cameras' FOVs should be secured so that camera handoff can be executed successfully and automatically before the object of interest becomes untraceable or unidentifiable. The proposed sensor planning method improves existing algorithms by adding handoff rate analysis, which preserves necessary overlapped FOVs for an optimal handoff success rate. In addition, special considerations such as resolution and frontal view requirements are addressed using two approaches: direct constraint and adaptive weight. The resulting camera placement is compared with a reference algorithm by Erdem and Sclaroff. Significantly improved handoff success rate and frontal view percentage are illustrated via experiments using typical office floor plans.

12. Visual Tracking Via Incremental Log-Euclidean Riemannian Subspace Learning

Xi Li, Weiming Hu, Zhongfei Zhang, Xiaoqin Zhang, Mingliang Zhu, Jian Cheng

Recently, a novel Log-Euclidean Riemannian metric [28] is proposed for statistics on symmetric positive definite (SPD) matrices. Under this metric, distances and Riemannian means take a much simpler form than the widely used affine-invariant Riemannian metric. Based on the Log-Euclidean Riemannian metric, we develop a tracking framework in this paper. In the framework, the covariance matrices of image features in the five modes are used to represent object appearance. Since a nonsingular covariance matrix is a SPD matrix lying on a connected Riemannian manifold, the Log-Euclidean Riemannian metric is used for statistics on the covariance matrices of image features. Further, we present an effective online Log-Euclidean Riemannian subspace learning algorithm which models the appearance changes of an object by incrementally learning a low-order Log-Euclidean eigenspace representation through adaptively updating the sample mean and eigenbasis. Tracking is then led by the Bayesian state inference framework in which a particle filter is used for propagating sample distributions over the time. Theoretic analysis and experimental evaluations demonstrate the promise and effectiveness of the proposed framework.

13. Increasing the Density of Active Appearance Models

Krishnan Ramnath, Simon Baker, Iain Matthews, Deva Ramanan

Active Appearance Models (AAMs) typically only use 50-100 mesh vertices because they are usually constructed from a set of training images with the vertices hand-labeled on them. In this paper, we propose an algorithm to increase the density of an AAM. Our algorithm operates by iteratively building the AAM, refitting the AAM to the training data, and refining the AAM. We compare our algorithm with the state of the art in optical flow algorithms and find it to be significantly more accurate. We also show that dense AAMs can be fit more robustly than sparse ones. Finally, we show how our algorithm can be used to construct AAMs automatically, starting with a single affine model that is subsequently refined to model non-planarity and non-rigidity.

14. 3D Ultrasound Tracking of The Left Ventricle Using One-Step Forward Prediction and Data Fusion of Collaborative Trackers

Lin Yang, Bogdan Georgescu, Yefeng Zheng, Peter Meer, Dorin Comaniciu

Tracking the left ventricle (LV) in 3D ultrasound data is a challenging task because of the poor image quality and speed requirements. Many previous algorithms applied standard 2D tracking methods to tackle the 3D problem. However, the performance is limited due to increased data size, landmarks ambiguity, signal drop-out or non-rigid deformation. In this paper we present a robust, fast and accurate 3D LV tracking algorithm. We propose a novel one-step forward prediction to generate the motion prior using motion manifold learning, and introduce two collaborative trackers to achieve both temporal consistency and failure recovery. Compared with tracking by detection and 3D optical flow, our algorithm provides the best results and subvoxel accuracy. The new tracking algorithm is completely automatic and computationally efficient. It requires less than 1.5 seconds to process a 3D volume which contains 4,925,440 voxels.

15. A Recursive Filter For Linear Systems on Riemannian Manifolds

Ambrish Tyagi, James Davis

We present an online, recursive filtering technique to model linear dynamical systems that operate on the state space of symmetric positive definite matrices (tensors) that lie on a Riemannian manifold. The proposed approach describes a predict-and-update computational paradigm, similar to a vector Kalman filter, to estimate the optimal tensor state. We adapt the original Kalman filtering algorithm to appropriately propagate the state over time and assimilate observations, while conforming to the geometry of the manifold. We validate our algorithm with synthetic data experiments and demonstrate its application to visual object tracking using covariance features.

16. Markerless Motion Capture of Man-Machine Interaction

Bodo Rosenhahn, Christian Schmaltz, Thomas Brox, Daniel Cremers, Joachim Weickert, Hans-Peter Seidel

This work deals with modeling and markerless tracking of athletes interacting with sports gear. In contrast to classical markerless tracking, the interaction with sports gear comes along with joint movement restrictions due to additional constraints: while humans can generally use all their joints, interaction with the equipment imposes a coupling between certain joints. A cyclist who performs a cycling pattern is one example: The feet are supposed to stay on the pedals, which are again restricted to move along a circular trajectory in 3D-space. In this paper, we present a markerless motion capture system that takes the lower-dimensional pose manifold into account by modeling the motion restrictions via soft constraints during pose optimization. Experiments with two different models, a cyclist and a snowboarder, demonstrate the applicability of the method. Moreover, we present motion capture results for challenging outdoor scenes including shadows and strong illumination changes.

17. Learning on Lie Groups for Invariant Detection and Tracking

Oncel Tuzel, Fatih Porikli, Peter Meer

This paper presents a novel learning based tracking model combined with object detection. The existing techniques proceed by linearizing the motion, which makes an implicit Euclidean space assumption. Most of the transformations used in computer vision have matrix Lie group structure. We learn the motion model on the Lie algebra and show that the formulation minimizes a first order approximation to the geodesic error. The learning model is extended to train a class specific tracking function, which is then integrated to an existing pose dependent object detector to build a pose invariant object detection algorithm. The proposed model can accurately detect objects in various poses, where the size of the search space is only a fraction compared to the existing object detection methods. The detection rate of the original detector is improved by more than 90% for large transformations.

18. Information-theoretic Active Scene Exploration

Eric Sommerlade, Ian Reid

Studies support the need for high resolution imagery to identify persons in surveillance videos. However, the use of telephoto lenses sacrifices a wider field of view and thereby increases the uncertainty of other, possibly more interesting events in the scene. Using zoom lenses offers the possibility of enjoying the benefits of both wide field of view and high resolution, but not simultaneously. We approach this problem of balancing these finite imaging resources -- or of exploration vs exploitation -- using an information-theoretic approach. We argue that the camera parameters -- pan, tilt and zoom -- should be set to maximise information gain, or equivalently minimising conditional entropy of the scene model, comprised of multiple targets and a yet unobserved one. The information content of the former is supplied directly by the uncertainties computed using a Kalman Filter tracker, while the latter is modelled using a “background” Poisson process whose parameters are learned from extended scene observations; together these yield an entropy for the scene. We support our argument with quantitative and qualitative analyses in simulated and real-world environments, demonstrating that this approach yields sensible exploration behaviours in which the camera alternates between obtaining close-up views of the targets while paying attention to the background, especially to areas of known high activity.

19. Parameterized Kernel Principal Component Analysis: Theory and Applications to Supervised and Unsupervised Image Alignment

Fernando De la Torre, Minh Hoai Nguyen

Parameterized Appearance Models (PAMs) (e.g., eigentracking, active appearance models, morphable models) use Principal Component Analysis (PCA) to model the shape and appearance of objects in images. Given a new image with an unknown appearance/shape configuration, PAMs can detect and track the object by optimizing the model's parameters that best match the image. While PAMs have numerous advantages for image registration relative to alternative approaches, they suffer from two major limitations: First, PCA cannot model non-linear structure in the data. Second, learning PAMs requires precise manually labeled training data. This paper proposes Parameterized Kernel Principal Component Analysis (PKPCA), an extension of PAMs that uses Kernel PCA (KPCA) for learning a non-linear appearance model invariant to rigid and/or non-rigid deformations. We demonstrate improved performance in supervised and unsupervised image registration, and present a novel application to improve the quality of manual landmarks in faces. In addition, we suggest a clean and effective matrix formulation for PKPCA.

20. Local Minima Free Parameterized Appearance Models

Minh Hoai Nguyen, Fernando De la Torre

Parameterized Appearance Models (PAMs) (e.g., Eigentracking, Active Appearance Models, Morphable Models) are commonly used to model the appearance and shape variation of objects in images. While PAMs have numerous advantages relative to alternate approaches, they have at least two drawbacks. First, they are especially prone to local minima in the fitting process. Second, often few if any of the local minima of the cost function correspond to acceptable solutions. To solve these problems, this paper proposes a method to learn a cost function by explicitly optimizing that the local minima occur at and only at the places corresponding to the correct fitting parameters. To the best of our knowledge, this is the first paper to address the problem of learning a cost function to explicitly model local properties of the error surface to fit PAMs. Synthetic and real examples show improvement in alignment performance in comparison with traditional approaches.

21. Super-Resolution from Image Sequence under Influence of Hot-Air Optical Turbulence

Masao Shimizu, Shin Yoshimura, Masayuki Tanaka, Masatoshi Okutomi

The appearance of a distant object, when viewed through a telephoto-lens, is often deformed nonuniformly by the influence of hot-air optical turbulence. The deformation is unsteady: an image sequence can include nonuniform movement of the object even if a stationary camera is used for a static object. This study proposes a multi-frame super-resolution reconstruction from such an image sequence. The process consists of the following three stages. In the first stage, an image frame without deformation is estimated from the sequence. However, there is little detailed information about the object. In the second stage, each frame in the sequence is aligned non-rigidly to the estimated image using a non-rigid deformation model. A stable non-rigid registration technique with a B-spline function is also proposed in this study for dealing with a textureless region. In the third stage, a multi-frame super-resolution reconstruction using the non-rigid deformation recovers the detailed information in the frame obtained in the first stage. Experiments using synthetic images demonstrate the accuracy and stability of the proposed non-rigid registration technique. Furthermore, experiments using real sequences underscore the effectiveness of the proposed process.

22. Calibration of an Articulated Camera System

Junzhou Chen, Kin Hong Wong

Multiple Camera Systems (MCS) have been widely used in many vision applications and attracted much attention recently. There are two principle types of MCS, one is the Rigid Multiple Camera System (RMCS); the other is the Articulated Camera System (ACS). In a RMCS, the relative poses (relative 3-D position and orientation) between the cameras are invariant. While, in an ACS, the cameras are articulated through movable joints, the relative pose between them may change. Therefore, through calibration of an ACS we want to find not only the relative poses between the cameras but also the positions of the joints in the ACS.

Although calibration methods for RMCS have been extensively developed during the past decades, the studies of ACS calibration are still rare. In this paper, two ACS calibration methods are proposed. The first one uses the feature correspondences between the cameras in the ACS. The second one requires only the ego-motion information of the cameras and can be used for the calibration of the non-overlapping view ACS. In both methods, the ACS is assumed to have performed general transformations in a static environment. The efficiency and robustness of the proposed methods are tested by simulation and real experiments. In the real experiment, the intrinsic and extrinsic parameters of the ACS are calibrated using the same image sequences, no extra data capturing step is required. The corresponding trajectory is recovered and illustrated using the calibration results of the ACS. To our knowledge, we are the first to study the calibration of ACS.

23. Recognizing Human Actions Using Multiple Features

Jingen Liu, Saad Ali, Mubarak Shah

In this paper, we propose a framework that fuses multiple features for improved action recognition in videos. The fusion of multiple features is important for recognizing actions as often a single feature based representation is not enough to capture the imaging variations (view-point, illumination etc.) and attributes of individuals (size, age, gender etc.). Hence, we use two types of features: i) a quantized vocabulary of local spatio-temporal (ST) volumes (or cuboids), and ii) a quantized vocabulary of spin-images, which aims to capture the shape deformation of the actor by considering actions as 3D objects (x, y, t). To optimally combine these features, we treat different features as nodes in a graph, where weighted edges between the nodes represent the strength of the relationship between entities. The graph is then embedded into a *k*-dimensional space subject to the criteria that similar nodes have Euclidian coordinates which are closer to each other. This is achieved by converting this constraint into a minimization problem whose solution is the eigenvectors of the graph Laplacian matrix. This procedure is known as Fiedler Embedding. The performance of the proposed framework is tested on publicly available data sets. The results demonstrate that fusion of multiple features helps in achieving improved performance, and allows retrieval of meaningful features and videos from the embedding space.

24. 3D-2D Spatiotemporal Registration for Sports Motion Analysis

Ruixuan Wang, Wee Kheng Leow, Hon Wai Leong

Computer systems are increasingly being used for sports training. Existing sports training systems either require expensive 3D motion capture systems or do not provide intelligent analysis of user's sports motion. This paper presents a framework for affordable and intelligent sports training systems for general users that require only single camera to record the user's motion. Sports motion analysis is formulated as a 3D-2D spatiotemporal motion registration problem. A novel algorithm is developed to perform spatiotemporal registration of the expert's 3D reference motion and a performer's 2D input video, thereby computing the deviation of the performer's motion from the expert's motion. The algorithm can effectively handle ambiguous situations in a single video such as depth ambiguity of body parts and partial occlusion. Test results show that, despite using only single video, the algorithm can compute 3D posture errors that reflect the performer's actual motion error.

25. Accurate Eye Center Location and Tracking Using Isophote Curvature

Roberto Valenti, Theo Gevers

The ubiquitous application of eye tracking is precluded by the requirement of dedicated and expensive hardware, such as infrared high definition cameras. Therefore, systems based solely on appearance (i.e., not involving active infrared illumination) are being proposed in literature. However, although these systems are able to successfully locate eyes, their accuracy is significantly lower than commercial eye tracking devices. Our aim is to perform very accurate eye center location and tracking, using a simple web cam. By means of a novel relevance mechanism, the proposed method makes use of isophote properties to gain invariance to linear lighting changes (contrast and brightness), to achieve rotational invariance and to keep low computational costs. In this paper we test our approach for accurate eye location and robustness to changes in illumination and pose, using the BioID and the Yale Face B databases, respectively. We demonstrate that our system can achieve a considerable improvement in accuracy over state of the art techniques.

26. Real-Time Pose Estimation of Articulated Objects using Low-Level Motion

Ben Daubney, David Gibson, Neill Campbell

We present a method that is capable of tracking and estimating pose of articulated objects in real-time. This is achieved by using a bottom-up approach to detect instances of the object in each frame, these detections are then linked together using a high-level a priori motion model. Unlike other approaches that rely on appearance, our method is entirely dependent on motion; initial low-level part detection is based on how a region moves as opposed to its appearance. This work is best described as Pictorial Structures using motion. A sparse cloud of points extracted using a standard feature tracker are used as observational data, this data contains noise that is not Gaussian in nature but systematic due to tracking errors. Using a probabilistic framework we are able to overcome both corrupt and missing data whilst still inferring new poses from a generative model. Our approach requires no manual initialisation and we show results for a number of complex scenes and different classes of articulated object, this demonstrates both the robustness and versatility of the presented technique.

27. Measuring camera translation by the dominant apical angle

Akihiko Torii, Michal Havlena, Tomáš Pajdla, Bastian Leibe

This paper provides a technique for measuring camera translation relatively w.r.t. the scene from two images. We demonstrate that the amount of the translation can be reliably measured for general as well as planar scenes by the most frequent apical angle, the angle under which the camera centers are seen from the perspective of the reconstructed scene points. Simulated experiments show that the dominant apical angle is a linear function of the length of the true camera translation. In a real experiment, we demonstrate that by skipping image pairs with too small motion, we can reliably initialize structure from motion, compute accurate camera trajectory in order to rectify images and use the ground plane constraint in recognition of pedestrians in a hand-held video sequence.

28. Fluid Flow Estimation Method Based on Physical Properties of Waves

Hidetomo Sakaino

This paper presents a fluid flow estimation method for ocean/river waves, clouds, and smoke based on the physical properties of waves. Most of the previous optical flow methods based on fluid dynamics/mechanics estimate a smooth flow using a continuity equation and/or div-curl velocity constraint. However, abrupt or inhomogeneous image motion changes in such fluid-like images are not estimated well. In this paper, we assume that many fluid-like motion changes are due to wave phenomena that lead to a brightness change. Thus, a wave generation equation is applied with a two-step optimization. A novel constraint based on the velocity-frequency relationship equation and a wave statistical property are used. The results of experiments on synthetic and real image sequences show the validity of our method.

29. Finding People in Archive Films through Tracking

Xiaofeng Ren

The goal of this work is to find all people in archive films. Challenges include low image quality, motion blur, partial occlusion, non-standard poses and crowded scenes. We base our approach on face detection and take a tracking/temporal approach to detection. Our tracker operates in two modes, following face detections whenever possible, switching to low-level tracking if face detection fails. With temporal correspondences established by tracking, we formulate detection as an inference problem in one-dimensional chains/tracks. We use a conditional random field model to integrate information across frames and to re-score tentative detections in tracks. Quantitative evaluations on full-length films show that the CRF-based temporal detector greatly improves face detection, increasing precision for about 30% (suppressing isolated false positives) and at the same time boosting recall for over 10% (recovering difficult cases where face detectors fail).

30. Discriminative Learning of Visual Words for 3D Human Pose Estimation

Huazhong Ning, Wei Xu, Yihong Gong, Thomas Huang

This paper addresses the problem of recovering 3D human pose from a single monocular image, using a discriminative bag-of-words approach. In previous work, the visual words are learned by unsupervised clustering algorithms. They capture the most common patterns and are good features for coarse-grain recognition tasks like object classification. But for those tasks which deal with subtle differences such as pose estimation, such representation may lack the needed discriminative power. In this paper, we propose to jointly learn the visual words and the pose regressors in a supervised manner. More specifically, we learn an individual distance metric for each visual word to optimize the pose estimation performance. The learned metrics rescale the visual words to suppress unimportant dimensions such as those corresponding to background. Another contribution is that we design an Appearance and Position Context (APC) local descriptor that achieves both selectivity and invariance while requiring no background subtraction. We test our approach on both a quasi-synthetic dataset and a real dataset (HumanEva) to verify its effectiveness. Our approach also achieves fast computational speed thanks to the integral histograms used in APC descriptor extraction and fast inference of pose regressors.

31. Modeling and Generating Complex Motion Blur for Real-time Tracking

Christopher Mei, Ian Reid

This article addresses the problem of real-time visual tracking in presence of complex motion blur. Previous authors have observed that efficient tracking can be obtained by matching blurred images instead of applying the computationally expensive task of deblurring [11]. The study was however limited to translational blur. In this work, we analyse the problem of tracking in presence of spatially variant motion blur generated by a planar template. We detail how to model the blur formation and parallelise the blur generation, enabling a real-time GPU implementation. Through the estimation of the camera exposure time, we discuss how tracking initialisation can be improved. Our algorithm is tested on challenging real data with complex motion blur where simple models fail. The benefit of blur estimation is shown for structure and motion.

32. Local Grouping for Optical Flow

Xiaofeng Ren

Optical flow estimation requires spatial integration, which essentially poses a grouping question: what points belong to the same motion and what do not. Classical local approaches to optical flow, such as Lucas-Kanade, use isotropic neighborhoods and have considerable difficulty near motion boundaries. In this work we utilize image-based grouping to facilitate spatial- and scale-adaptive integration. We define soft spatial support using pairwise affinities computed through intervening contour. We sample images at edges and corners, and iteratively estimate affine motion at sample points. Figure-ground organization further improves grouping and flow estimation near boundaries. We show that affinity-based spatial integration enables reliable flow estimation and avoids erroneous motion propagation from and/or across object boundaries. We demonstrate our approach on the Middlebury flow dataset.

33. Motion blur identification from image gradients

Hui Ji, Chaoqiang Liu

Restoration of a degraded image from motion blurring is highly dependent on the estimation of the blurring kernel. Most of the existing motion deblurring techniques model the blurring kernel with a shift-invariant box filter, which holds true only if the motion among images is of uniform velocity. In this paper, we present a spectral analysis of image gradients, which leads to a better configuration for identifying the blurring kernel of more general motion types (uniform velocity motion, accelerated motion and vibration). Furthermore, we introduce a hybrid Fourier-Radon transform to estimate the parameters of the blurring kernel with improved robustness to noise over available techniques. The experiments on both simulated images and real images show that our algorithm is capable of accurately identifying the blurring kernel for a wider range of motion types.

34. Articulated Shape Matching Using Laplacian Eigenfunctions and Unsupervised Point Registration

Diana Mateus, Radu Horaud, David Knossow, Fabio Cuzzolin, Edmond Boyer

Matching articulated shapes represented by voxel-sets reduces to maximal sub-graph isomorphism when each set is described by a weighted graph. Spectral graph theory can be used to map these graphs onto lower dimensional spaces and match shapes by aligning their embeddings in virtue of their invariance to change of pose. Classical graph isomorphism schemes relying on the ordering of the eigenvalues to align the eigenspaces fail when handling large data-sets or noisy data. We derive a new formulation that finds the best alignment between two congruent K-dimensional sets of points by selecting the best subset of eigenfunctions of the Laplacian matrix. The selection is done by matching eigenfunction signatures built with histograms, and the retained set provides a smart initialization for the alignment problem with a considerable impact on the overall performance. Dense shape matching casted into graph matching reduces then, to point registration of embeddings under orthogonal transformations; the registration is solved using the framework of unsupervised clustering and the EM algorithm. Maximal subset matching of non identical shapes is handled by defining an appropriate outlier class. Experimental results on challenging examples show how the algorithm naturally treats changes of topology, shape variations and different sampling densities.

35. Homography Based Multiple Camera Detection and Tracking of People in a Dense Crowd

Ran Eshel, Yael Moses

Tracking people in a dense crowd is a challenging problem for a single camera tracker due to occlusions and extensive motion that make human segmentation difficult. In this paper we suggest a method for simultaneously tracking all the people in a densely crowded scene using a set of cameras with overlapping fields of view. To overcome occlusions, the cameras are placed at a high elevation and only people's heads are tracked. Head detection is still difficult since each foreground region may consist of multiple subjects. By combining data from several views, height information is extracted and used for head segmentation. The head tops, which are regarded as 2D patches at various heights, are detected by applying intensity correlation to aligned frames from the different cameras. The detected head tops are then tracked using common assumptions on motion direction and velocity. The method was tested on sequences in indoor and outdoor environments under challenging illumination conditions. It was successful in tracking up to 21 people walking in a small area (2.5 people per m^2), in spite of severe and persistent occlusions.

36. Adaptive and Constrained Algorithms for Inverse Compositional Active Appearance Model Fitting

George Papandreou, Petros Maragos

Parametric models of shape and texture such as Active Appearance Models (AAMs) are diverse tools for deformable object appearance modeling and have found important applications in both image synthesis and analysis problems. Among the numerous algorithms that have been proposed for AAM fitting, those based on the inverse-compositional image alignment technique have recently received considerable attention due to their potential for high efficiency. However, existing fitting algorithms perform poorly when used in conjunction with models exhibiting significant appearance variation, such as AAMs trained on multiple-subject human face images. We introduce two enhancements to inverse-compositional AAM matching algorithms in order to overcome this limitation. First, we propose fitting algorithm **adaptation**, by means of (a) fitting matrix adjustment and (b) AAM mean template update. Second, we show how prior information can be incorporated and **constrain** the AAM fitting process. The inverse-compositional nature of the algorithm allows efficient implementation of these enhancements. Both techniques substantially improve AAM fitting performance, as demonstrated with experiments on publicly available multi-person face datasets.

37. An Integrated Background Model for Video Surveillance Based on Primal Sketch and 3D Scene Geometry

Hu Wenze, Gong Haifeng, Zhu Song-Chun

This paper presents a novel integrated background model for video surveillance. Our model uses a primal sketch representation for image appearance and 3D scene geometry to capture the ground plane and major surfaces in the scene. The primal sketch model divides the background image into three types of regions -- flat, sketchable and textured. The three types of regions are modeled respectively by mixture of Gaussians, image primitives and LBP histograms. We calibrate the camera and recover important planes such as ground, horizontal surfaces, walls, stairs in the 3D scene, and use geometric information to predict the sizes and locations of foreground blobs to further reduce false alarms. Compared with the state-of-the-art background modeling methods, our approach is more effective, especially for indoor scenes where shadows, highlights and reflections of moving objects and camera exposure adjusting usually cause problems. Experimental results demonstrate that our approach improves the performance of background/foreground separation at pixel level, and the integrated video surveillance system at the object and trajectory level.

38. Object Tracking and Detection after Occlusion via Numerical Hybrid Local and Global Mode-seeking

Zhaozheng Yin, Robert Collins

Given an object model and a black-box measure of similarity between the model and candidate targets, we consider visual object tracking as a numerical optimization problem. During normal tracking conditions when the object is visible from frame to frame, local optimization is used to track the local mode of the similarity measure in a parameter space of translation, rotation and scale. However, when the object becomes partially or totally occluded, such local tracking is prone to failure, especially when common prediction techniques like the Kalman filter do not provide a good estimate of object parameters in future frames. To recover from these inevitable tracking failures, we consider object detection as a global optimization problem and solve it via Adaptive Simulated Annealing (ASA), a method that avoids becoming trapped at local modes and is much faster than exhaustive search. As a Monte Carlo approach, ASA stochastically samples the parameter space, in contrast to local deterministic search. We apply cluster analysis on the sampled parameter space to redetect the object and renew the local tracker. Our numerical hybrid local and global mode-seeking tracker is validated on challenging airborne videos with heavy occlusion and large camera motions. Our approach outperforms state-of-the-art trackers on the VIVID benchmark datasets.

39. A Loopy Belief Propagation Approach for Robust Background Estimation

Xun Xu, Thomas S. Huang

Background estimation, i.e., automatic recovery of the background image from a sequence of images containing moving foreground objects, is an important module in many applications, e.g., surveillance and video segmentation. In this paper, we present a simple, yet effective and robust approach for background estimation based on Loopy Belief Propagation. Robustness of the proposed approach means: (i) minimal assumption on the input frames, and (ii) no need to tune parameters. Basically, the background can be recovered even when the occluding foreground objects stay still for a long time. Furthermore, no motion information needs to be known or estimated for the foreground objects, which implies that background can be recovered from a set of frames which are not consecutive temporally. Analysis and experiments are provided to compare the proposed approach to related methods. Experimental results on typical surveillance videos demonstrate the effectiveness of our approach.

40. Layered graphical models for tracking partially-occluded objects

Vitaly Ablavsky, Ashwin Thangali, Stan Sclaroff

Partial occlusions are commonplace in a variety of real world computer vision applications: surveillance, intelligent environments, assistive robotics, autonomous navigation, etc. While occlusion handling methods have been proposed, most methods tend to break down when confronted with numerous occluders in a scene. In this paper, a layered image-plane representation for tracking people through substantial occlusions is proposed. An image-plane representation of motion around an object is associated with a pre-computed graphical model, which can be instantiated efficiently during online tracking. A global state and observation space is obtained by linking transitions between layers. A Reversible Jump Markov Chain Monte Carlo approach is used to infer the number of people and track them online. The method outperforms two state-of-the-art methods for tracking over extended occlusions, given videos of a parking lot with numerous vehicles and a laboratory with many desks and workstations.

41. Linear approach to motion estimation using generalized camera models

Hongdong Li, Richard Hartley, Jae-hak Kim

A well-known theoretical result for motion estimation using the generalized camera model is that 17 corresponding image rays can be used to solve linearly for the motion of a generalized camera. However, this paper shows that for many common configurations of the generalized camera models (e.g., multi-camera rig, catadioptric camera etc.), such a simple 17-point algorithm does not exist, due to some previously overlooked ambiguities.

We further discover that, despite the above ambiguities, we are still able to solve the motion estimation problem effectively by a new algorithm proposed in this paper. Our algorithm is essentially linear, easy to implement, and the computational efficiency is very high. Experiments on both real and simulated data show that the new algorithm achieves reasonably high accuracy as well.

42. The Kneed Walker for Human Pose Tracking

Marcus A. Brubaker, David J. Fleet

The Kneed Walker is a physics-based model derived from a planar biomechanical characterization of human locomotion. By controlling torques at the knees, hips and torso, the model captures a full range of walking motions with foot contact and balance. Constraints are used to properly handle ground collisions and joint limits. A prior density over walking motions is based on dynamics that are optimized for efficient cyclic gaits over a wide range of natural human walking speeds and step lengths, on different slopes. The generative model used for monocular tracking comprises the Kneed Walker prior, a 3D kinematic model constrained to be consistent with the underlying dynamics, and a simple measurement model in terms of appearance and optical flow. The tracker is applied to people walking with varying speeds, on hills, and with occlusion.

43. Extracting a Fluid Dynamic Texture and the Background from Video

Bernard Ghanem, Narendra Ahuja

Given the video of a still background occluded by a fluid dynamic texture (FDT), this paper addresses the problem of separating the video sequence into its two constituent layers. One layer corresponds to the video of the unoccluded background, and the other to that of the dynamic texture, as it would appear if viewed against a black background. The model of the dynamic texture is unknown except that it represents fluid flow. We present an approach that uses the image motion information to simultaneously obtain a model of the dynamic texture and separate it from the background which is required to be still. Previous methods have considered occluding layers whose dynamics follows simple motion models (e.g., periodic or 2D parametric motion). FDTs considered in this paper exhibit complex stochastic motion. We consider videos showing an FDT layer (e.g., pummeling smoke or heavy rain) in front of a static background layer (e.g., brick building). We propose a novel method for simultaneously separating these two layers and learning a model for the FDT. Due to the fluid nature of the DT, we are required to learn a model for both the spatial appearance and the temporal variations (due to changes in density) of the FDT, along with a valid estimate of the background. We model the frames of a sequence as being produced by a continuous HMM, characterized by transition probabilities based on the Navier-Stokes equations for fluid dynamics, and by generation probabilities based on the convex matting of the FDT with the background. We learn the FDT appearance, the FDT temporal variations, and the background by maximizing their joint probability using Interactive Conditional Modes (ICM). Since the learned model is generative, it can be used to synthesize new videos with different backgrounds and density variations. Experiments on videos that we compiled demonstrate the performance of our method.

44. Fast Track Matching and Event Detection

Tao Ding, Mario Sznaier, Octavia I. Camps

This paper addresses the problems of track stitching and dynamic event detection in a sequence of frames. The input data consists of tracks, possibly fragmented due to occlusion, belonging to multiple targets. The goals are to (i) establish track identity across occlusion, and (ii) detect points where the motion of these targets undergo substantial changes. The main result of the paper is a simple, computationally inexpensive approach that achieves these goals in a unified way. Given a continuous track, the main idea is to detect changes in the dynamics by parsing it into segments according to the complexity of the model required to explain the observed data. Intuitively, changes in this complexity correspond to points where the dynamics change. Since the problem of estimating the complexity of the underlying model can be reduced to estimating the rank of a matrix constructed from the observed data, these changes can be found with a simple algorithm, computationally no more expensive that a sequence of SVDs. Proceeding along the same lines, fragmented tracks corresponding to multiple targets can be linked by searching for sets corresponding to minimal complexity joint models. As we show in the paper, this problem can be reduced to a semi-definite optimization and efficiently solved with commonly available software.

45. Tracking Rotating Fluids in Realtime using Snapshots

Sai Ravela, John Marshall, Christopher Hill, Andrew Wong, Scott Stransky

We present a model-based system for tracking rotating fluids, and apply it to a laboratory study of atmospheric circulation. Tracking is accomplished by filtering uncertain and high-dimensional states of a nonlinear general circulation model with optical measurements of the physical fluid's velocity. Realtime performance is achieved by using a nonuniform discretization of the model's spatial resolution, and by using time-snapshots of model-state to construct spatially-localized reduced-rank square-root representations of forecast uncertainty.

Realtime performance, economical and repeatable experimentation, and a direct connection to planetary flows implies that the proposed physical-numerical coupling can be useful for addressing many perceptual geophysical fluid dynamics problems. To the best of our knowledge, such a system has not hitherto been reported.

46. Granularity and Elasticity Adaptation in Visual Tracking

Ming Yang, Ying Wu

The observation models in tracking algorithms are critical to both tracking performance and applicable scenarios but are often simplified to focus on fixed level of certain target properties such as appearances and structures. In this paper, we propose a unified tracking paradigm in which targets are represented by Markov random fields of interest regions and introduce a new way to adapt observation models by automatically tuning the feature granularity and model elasticity, i.e., the abstraction level of features and the model's degree of flexibility to tolerate deformations. Specifically, we employ a multi-scale scheme to extract features from interest regions and adjust the parameters of the potential functions of the MRF model to maximize the likelihoods of tracking results. Experiments demonstrate the method can estimate translation, scaling and rotation and deal with deformation, partial occlusions, and camouflage objects within this unified framework.

47. Real Time Object Tracking based on Dynamic Feature Grouping with Background Subtraction

ZuWhan Kim

Object detection and tracking has various application areas including intelligent transportation systems. We introduce an object detection and tracking approach that combines the background subtraction algorithm and the feature tracking and grouping algorithm. We first present an augmented background subtraction algorithm which uses a low-level feature tracking as a cue. The resulting background subtraction cues are used to improve the feature detection and grouping result. We then present a dynamic multi-level feature grouping approach that can be used in real time applications and also provides high-quality trajectories. Experimental results from video clips of a challenging transportation application are presented.

48. Learning the Viewpoint Manifold for Action Recognition

Richard Souvenir, Justin Babbs

Researchers are increasingly interested in providing video-based, view-invariant action recognition for human motion. Addressing this problem will lead to more accurate modeling and analysis of the type of unconstrained video commonly collected in the areas of athletics and medicine. Previous viewpoint-invariant methods use multiple cameras in both the training and testing phases of action recognition or require storing many examples of a single action from multiple viewpoints. In this paper, we present a framework for learning a compact representation of primitive actions (e.g., walk, punch, kick, sit) that can be used for video obtained from a single camera for simultaneous action recognition and viewpoint estimation. Using our method, which models the low-dimensional structure of these actions relative to viewpoint, we show recognition rates on a publicly available data set previously only acheieved using multiple simultaneous views.

49. Optical Flow Estimation using Fourier Mellin Transform

Huy Tho Ho, Roland Goecke

In this paper, we propose a novel method of computing the optical flow using the Fourier Mellin Transform (FMT). Each image in a sequence is divided into a regular grid of patches and the optical flow is estimated by calculating the phase correlation of each pair of co-sited patches using the FMT. By applying the FMT in calculating the phase correlation, we are able to estimate not only the pure translation, as limited in the case of the basic phase correlation techniques, but also the scale and rotation motion of image patches, i.e., full similarity transforms. Moreover, the motion parameters of each patch can be estimated to subpixel accuracy based on a recently proposed algorithm that uses a 2D esinc function in fitting the data from the phase correlation output. We also improve the estimation of the optical flow by presenting a method of smoothing the field by using a vector weighted average filter. Finally, experimental results, using publicly available data sets are presented, demonstrating the accuracy and improvements of our method over previous optical flow methods.

50. Moving Shape Dynamics: A Signal Processing Perspective

Liang Wang, Xin Geng, Christopher Leckie, Ramamohanarao Kotagiri

This paper provides a new perspective on human motion analysis, namely regarding human motions in video as general discrete time signals. While this seems an intuitive idea, research on human motion analysis has attracted little attention from the signal processing community. Sophisticated signal processing techniques create important opportunities for new solutions to the problem of human motion analysis. This paper investigates how the deformations of human silhouettes (or shapes) during articulated motion can be used as discriminating features to implicitly capture motion dynamics. In particular, we demonstrate the applicability of two widely used signal transform methods, namely the Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT), for characterization and recognition of human motion sequences. Experimental results show the effectiveness of the proposed method on two state-of-the-art data sets.

51. A Theoretical Analysis of Linear and Multi-linear Models of Image Appearance

Yilei Xu, Amit Roy-Chowdhury

Linear and multi-linear models of object shape/appearance (PCA, 3DMM, AAM/ASM, multilinear tensors) have been very popular in computer vision. In this paper, we analyze the validity of these models from the fundamental physical laws of object motion and image formation. We rigorously prove that the image appearance space can be closely approximated to be locally multilinear, with the illumination subspace being bilinearly combined with the direct sum of the motion, deformation and texture subspaces. This result allows us to understand theoretically many of the successes and limitations of the linear and multi-linear approaches existing in the computer vision literature, and also identifies some of the conditions under which they are valid. It provides an analytical representation of the image space in terms of different physical factors that affect the image formation process. Experimental analysis of the accuracy of the theoretical models is performed as well as tracking on real data using the analytically derived basis functions of this space.

52. Boosting Adaptive Linear Weak Classifiers for Online learning and Tracking

Toufiq Parag, Fatih Porikli, Ahmed Elgammal

Online boosting methods have recently been used successfully for tracking, background subtraction etc. Conventional online boosting algorithms emphasize on interchanging new weak classifiers/features to adapt with the change over time. We are proposing a new online boosting algorithm where the form of the weak classifiers themselves are modified to cope with scene changes. Instead of replacement, the parameters of the weak classifiers are altered in accordance with the new data subset presented to the online boosting process at each time step. Thus we may avoid altogether the issue of how many weak classifiers to be replaced to capture the change in the data or which efficient search algorithm to use for a fast retrieval of weak classifiers. A computationally efficient method has been used in this paper for the adaptation of linear weak classifiers. The proposed algorithm has been implemented to be used both as an online learning and a tracking method. We show quantitative and qualitative results on both UCI datasets and several video sequences to demonstrate improved performance of our algorithm.

53. Discriminative Human Action Segmentation and Recognition using Semi-Markov Model

Qinfeng Shi, Li Wang, Li Cheng, Alex Smola

Given an input video sequence of one person conducting a sequence of continuous actions, we consider the problem of jointly segmenting and recognizing actions. We propose a discriminative approach to this problem under a semi-Markov model framework, where we are able to define a set of features over input-output space that captures the characteristics on boundary frames, action segments and neighboring action segments, respectively. In addition, we show that this method can also be used to recognize the person who performs in this video sequence. A Viterbi-like algorithm is devised to help efficiently solve the induced optimization problem. Experiments on a variety of datasets demonstrate the effectiveness of the proposed method.

54. Drift-free Tracking of Rigid and Articulated Objects

Juergen Gall, Bodo Rosenhahn, Hans-Peter Seidel

Model-based 3D tracker estimate the position, rotation, and joint angles of a given model from video data of one or multiple cameras. They often rely on image features that are tracked over time but the accumulation of small errors results in a drift away from the target object. In this work, we address the drift problem for the challenging task of human motion capture and tracking in the presence of multiple moving objects where the error accumulation becomes even more problematic due to occlusions. To this end, we propose an analysis-by-synthesis framework for articulated models. It combines the complementary concepts of patch-based and region-based matching to track both structured and homogeneous body parts. The performance of our method is demonstrated for rigid bodies, body parts, and full human bodies where the sequences contain fast movements, self-occlusions, multiple moving objects, and clutter. We also provide a quantitative error analysis and comparison with other model-based approaches.

55. Learning stick-figure models using nonparametric Bayesian priors over trees

Edward W. Meeds, David A. Ross, Richard S. Zemel, Sam T. Roweis

We present a probabilistic stick-figure model that uses a nonparametric Bayesian distribution over trees for its structure prior. Sticks are represented by nodes in a tree in such a way that their parameter distributions are probabilistically centered around their parent node. This prior enables the inference procedures to learn multiple explanations for motion-capture data, each of which could be trees of different depth and path lengths. Thus, the algorithm can automatically determine a reasonable distribution over the number of sticks in a given dataset and their hierarchical relationships. We provide experimental results on several motion-capture datasets, demonstrating the model's ability to recover plausible stick-figure structure, and also the model's robust behavior when faced with occlusion.

56. Distributed Data Association and Filtering for Multiple Target Tracking

Ting Yu, Ying Wu, Nils O. Krahnstoever, Peter H. Tu

This paper presents a novel distributed framework for multi-target tracking with an efficient data association computation. A decentralized representation of trackers' motion and association variables is adopted. Considering the interleaved nature of data association and tracker filtering, the multi-target tracking is formulated as a missing data problem, and the solution is found by the proposed variational EM algorithm. We analytically show that 1) the posteriori distributions of trackers' motions (the real interests in terms of tracking applications) can be effectively computed in the E-step of the EM iterations, and 2) the solution of trackers' association variables can be pursued under a derived graph-based discrete optimization formulation, thus efficiently estimated in the M-step by the recently emerging graph optimization algorithms. The proposed approach is very general such that sophisticated data association priori and likelihood function can be easily incorporated. This general framework is tested with both simulation data and real world surveillance video. The reported qualitative and quantitative studies verify the effectiveness and low computational cost of the algorithm.

57. Modeling complex luminance variations for target tracking

Christophe Collewet, Eric Marchand

Lambert's model is widely used in low level computer vision algorithms such as matching, tracking or optical flow computation for example. However, it is well known that these algorithms often fail when they face complex luminance variations. Therefore, we revise in this paper the underlying hypothesis of its temporal constancy and propose a new optical flow constraint. To do that, we use the Blinn-Phong reflection model to take into account that the scene may move with respect to the lighting and/or to the observer, and that specular highlights may occur. To validate in practice these analytical results, we consider the case where a camera is mounted on a robot end-effector with a lighting mounted on this camera and show experimental results of target tracking by visual servoing. Such an approach requires to analytically compute the luminance variations due to the observer motion which can be easily derived from our revised optical flow constraint. In addition, while the visual servoing classical approaches rely on geometric features, we present here a new method that directly relies on the luminance of all pixels in the image which does not require any tracking or matching process.

58. Optical Flow Estimation with Uncertainties through Dynamic MRFs

Ben Glocker, Nikos Paragios, Nikos Komodakis, Georgios Tziritas, Nassir Navab

In this paper, we propose a novel dynamic discrete framework to address image morphing with application to optical flow estimation. We reformulate the problem using a number of discrete displacements, and therefore the estimation of the morphing parameters becomes a tractable matching criteria independent combinatorial problem which is solved through the FastPD algorithm. In order to overcome the main limitation of discrete approaches (low dimensionality of the label space is unable to capture the continuous nature of the expected solution), we introduce a dynamic behavior in the model where the plausible discrete deformations (displacements) are varying in space (across the domain) and time (different states of the process - successive morphing states) according to the local uncertainty of the obtained solution.

59. Boosted Deformable Model for Human Body Alignment

Xiaoming Liu, Ting Yu, Thomas Sebastian, Peter Tu

This paper studies image alignment, the problem of learning a shape and appearance model from labeled data and efficiently fitting the model to a non-rigid object with large variations. Given a set of images with manually labeled landmarks, our model representation consists of a shape component represented by a Point Distribution Model and an appearance component represented by a collection of local features, trained discriminatively as a two-class classifier using boosting. Images with ground truth landmarks are the positive training samples while those with perturbed landmarks are considered as negatives. Enabled by piece-wise affine warping, corresponding local feature positions across all training samples form a hypothesis space for boosting. Image alignment is performed by maximizing the boosted classifier score, which is our distance measure, through iteratively mapping the feature positions to the image, and computing the gradient direction of the score with respect to the shape parameter. We apply this approach to human body alignment from surveillance-type images. We conduct experiments on the MIT pedestrian database where the body size is approximately 110 × 46 pixels, and demonstrate our real-time alignment capability.

60. Enforcing Non-Positive Weights for Stable Support Vector Tracking

Simon Lucey

In this paper we demonstrate that the support vector tracking (SVT) framework first proposed by Avidan is equivalent to the canonical Lucas-Kanade (LK) algorithm with a weighted Euclidean norm. From this equivalence we empirically demonstrate that in many circumstances the canonical SVT approach is unstable, and characterize these circumstances theoretically. We then propose a novel “nonpositive support kernel machine” (NSKM) to circumvent this limitation and allow the effective use of discriminative classification within the weighted LK framework. This approach ensures that the pseudo-Hessian realized within the weighted LK algorithm is positive semidefinite which allows for fast convergence and accurate alignment/tracking. A further benefit of our proposed method is that the NSKM solution results in a much sparser kernel machine than the canonical SVM leading to sizeable computational savings and much improved alignment performance.

61. Meshless Deformable Models for LV Motion Analysis

Xiaoxu Wang, Ting Chen, Dimitris Metaxas, Leon Axel

We propose a novel meshless deformable model for in vivo cardiac left ventricle (LV) 3D motion estimation. As a relatively new technology, tagged MRI (tMRI) provides a direct and noninvasive way to reveal local deformation of the myocardium, which creates a large amount of heart motion data which requiring quantitative analysis. In our study, we sample the heart motion sparsely at intersections of three sets of orthogonal tagging planes and then use a new meshless deformable model to recover the dense 3D motion of the myocardium temporally during the cardiac cycle. We compute external forces at tag intersections based on tracked local motion and redistribute the force to meshless particles throughout the myocardium. Internal constraint forces at particles are derived from local strain energy using a Moving Least Squares (MLS) method. The dense 3D motion field is then computed and updated using the Lagrange equation. The new model avoids the singularity problem of mesh-based models and is capable of tracking large deformation with high efficiency and accuracy. In particular, the model performs well even when the control points (tag intersections) are relatively sparse. We tested the performance of the meshless model on a numerical phantom, as well as in vivo heart data of healthy subjects and patients. The experimental results show that the meshless deformable model can fully recover the myocardium motion in 3D.

62. 3D Tracking of Shoes for Virtual Mirror Applications

Peter Eisert, Philipp Fechteler, Juergen Rurainsky

In this paper, augmented reality techniques are used in order to create a Virtual Mirror for the real-time visualization of customized sports shoes. Similar to looking into a mirror when trying on new shoes in a shop, we create the same impression but for virtual shoes that the customer can design individually. For that purpose, we replace the real mirror by a large display that shows the mirrored input of a camera capturing the legs and shoes of a person. 3-D Tracking of both feet and exchanging the real shoes by computer graphics models gives the impression of actually wearing the virtual shoes. The 3-D motion tracker presented in this paper, exploits mainly silhouette information to achieve robust estimates for both shoes from a single camera view. The use of a hierarchical approach in an image pyramid enables real-time estimation at frame rates of more than 30 frames per second.

63. On Handling Uncertainty in the Fundamental Matrix for Scene and Motion Adaptive Pose Recovery

Sreenivas Sukumar, Hamparsum Bozdogan, David Page, Andreas Koschan, Abidi Mongi

The estimation of the fundamental matrix is the key step in feature-based camera ego-motion estimation for applications in scene modeling and vehicle navigation. In this paper, we present a new method of analyzing and further reducing the risk in the fundamental matrix due to the choice of a particular feature detector, the choice of the matching algorithm, the motion model, iterative hypothesis generation and verification paradigms. Our scheme makes use of model-selection theory to guide the switch to optimal methods for fundamental matrix estimation within the hypothesis-and-test architecture. We demonstrate our proposed method for vision-based robot localization in large-scale environments where the environment is constantly changing and navigation within the environment is unpredictable.

64. Observe-and-Explain: A New Approach for Multiple Hypotheses Tracking of Humans and Objects

Michael S. Ryoo, J. K. Aggarwal

This paper presents a novel approach for tracking humans and objects under severe occlusion. We introduce a new paradigm for multiple hypotheses tracking, observe-and-explain, as opposed to the previous paradigm of hypothesize-and-test. Our approach efficiently enumerates multiple possibilities of tracking by generating several likely `explanations' after concatenating a sufficient amount of observations. The computational advantages of our approach over the previous paradigm under severe occlusions are presented. The tracking system is implemented and tested using the i-Lids dataset, which consists of videos of humans and objects moving in a London subway station. The experimental results show that our new approach is able to track humans and objects accurately and reliably even when they are completely occluded, illustrating its advantage over previous approaches.

65. Privacy Preserving Crowd Monitoring: Counting People without People Models or Tracking

Antoni B. Chan, Zhang-Sheng John Liang, Nuno Vasconcelos

We present a privacy-preserving system for estimating the size of inhomogeneous crowds, composed of pedestrians that travel in different directions, without using explicit object segmentation or tracking. First, the crowd is segmented into components of homogeneous motion, using the mixture of dynamic textures motion model. Second, a set of simple holistic features is extracted from each segmented region, and the correspondence between features and the number of people per segment is learned with Gaussian Process regression. We validate both the crowd segmentation algorithm, and the crowd counting system, on a large pedestrian dataset (2000 frames of video, containing 49,885 total pedestrian instances). Finally, we present results of the system running on a full hour of video.

66. Fuzzy Chamfer Distance and its Probabilistic Formulation for Visual Tracking

Yonggang Jin, Farzin Mokhtarian, Miroslaw Bober, John Illingworth

The paper presents a fuzzy chamfer distance and its probabilistic formulation for edge-based visual tracking. First, connections of the chamfer distance and the Hausdorff distance with fuzzy objective functions for clustering are shown using a reformulation theorem. A fuzzy chamfer distance (FCD) based on fuzzy objective functions and a probabilistic formulation of the fuzzy chamfer distance (PFCD) based on data association methods are then presented for tracking, which can all be regarded as reformulated fuzzy objective functions and minimized with iterative algorithms. Results on challenging sequences demonstrate the performance of the proposed tracking method.

67. On Errors-In-Variables Regression with Arbitrary Covariance and its Application to Optical Flow Estimation

Björn Andres, Claudia Kondermann, Daniel Kondermann, Ullrich Köthe, Fred A. Hamprecht, Christoph S. Garbe

Linear inverse problems in computer vision, including motion estimation, shape fitting and image reconstruction, give rise to parameter estimation problems with highly correlated errors in variables. Established total least squares methods estimate the most likely corrections $\hat{A}$ and $\hat{b}$ to a given data matrix $[A, b]$ perturbed by additive Gaussian noise, such that there exists a solution y with $[A + \hat{A}\ , b + \hat{b}\ ]y = 0$. In practice, regression imposes a more restrictive constraint namely the existence of a solution x with $[A + \hat{A}]x = [b + \hat{b}\ ]$. In addition, more complicated correlations arise canonically from the use of linear filters. We, therefore, propose a maximum likelihood estimator for regression in the general case of arbitrary positive definite covariance matrices. We show that $\hat{A}$, $\hat{b}$ and x can be found simultaneously by the unconstrained minimization of a multivariate polynomial which can, in principle, be carried out by means of a Gröbner basis. Results for plane fitting and optical flow computation indicate the superiority of the proposed method.

68. Face Tracking and Recognition with Visual Constraints in Real-World Videos

Minyoung Kim, Sanjiv Kumar, Vladimir Pavlovic, Henry Rowley

We address the problem of tracking and recognizing faces in real-world, noisy videos. We track faces using a tracker that adaptively builds a target model reflecting changes in appearance, typical of a video setting. However, adaptive appearance trackers often suffer from drift, a gradual adaptation of the tracker to non-targets. To alleviate this problem, our tracker introduces visual constraints using a combination of generative and discriminative models in a particle filtering framework. The generative term conforms the particles to the space of generic face poses while the discriminative one ensures rejection of poorly aligned targets. This leads to a tracker that significantly improves robustness against abrupt appearance changes and occlusions, critical for the subsequent recognition phase. Identity of the tracked subject is established by fusing pose-discriminant and person-discriminant features over the duration of a video sequence. This leads to a robust video-based face recognizer with state-of-the-art recognition performance. We test the quality of tracking and face recognition on realworld noisy videos from YouTube as well as the standard Honda/UCSD database. Our approach produces successful face tracking results on over 80% of all videos without video or person-specific parameter tuning. The good tracking performance induces similarly high recognition rates: 100% on Honda/UCSD and over 70% on the YouTube set containing 35 celebrities in 1500 sequences.

69. Least Squares Congealing for Unsupervised Alignment of Images

Mark Cox, Simon Lucey, Jeffrey Cohn, Sridha Sridharan

In this paper, we present an approach we refer to as “least squares congealing” which provides a solution to the problem of aligning an ensemble of images in an unsupervised manner. Our approach circumvents many of the limitations existing in the canonical “congealing” algorithm. Specifically, we present an algorithm that:- (i) is able to simultaneously, rather than sequentially, estimate warp parameter updates, (ii) exhibits fast convergence and (iii) requires no pre-defined step size. We present alignment results which show an improvement in performance for the removal of unwanted spatial variation when compared with the related work of Learned-Miller on two datasets, the MNIST hand written digit database and the MultiPIE face database.

70. A Hybrid Camera for Motion Deblurring and Depth Map Super-Resolution

Feng Li, Jingyi Yu, Jinxiang Chai

We present a hybrid camera that combines the advantages of a high resolution camera and a high speed camera. Our hybrid camera consists of a pair of low-resolution high-speed (LRHS) cameras and a single high-resolution low-speed (HRLS) camera. The LRHS cameras are able to capture fast-motion with little motion blur. They also form a stereo pair and provide a low-resolution depth map. The HRLS camera provides a high spatial resolution but also introduces severe motion blur when capturing fast moving objects. We develop efficient algorithms to simultaneously motion-deblur the HRLS image and reconstruct a high resolution depth map. Our method estimates the motion flow in the LRHS pair and then warps the flow field to the HRLS camera to estimate the point spread function (PSF). We then deblur the HRLS image and use the resulting image to enhance the low-resolution depth map using joint bilateral filters. We demonstrate the hybrid camera in depth map super-resolution and motion deblurring with spatially varying kernels. Experiments show that our framework is robust and highly effective.

71. Visual Tracking with Histograms and Articulating Blocks

S. M. Shahed Nejhum, Jeffrey Ho, Ming-Hsuan Yang

We propose an algorithm for accurate tracking of (articulated) objects using online update of appearance and shape. The challenge here is to model foreground appearance with histograms in a way that is both efficient and accurate. In this algorithm, the constantly changing foreground shape is modeled as a small number of rectangular blocks, whose positions within the tracking window are adaptively determined. Under the general assumption of stationary foreground appearance, we show that robust object tracking is possible by adaptively adjusting the locations of these blocks. Implemented in MATLAB without substantial optimization, our tracker runs already at 3.7 frames per second on a 3GHz machine. Experimental results have demonstrated that the algorithm is able to efficiently track articulated objects undergoing large variation in appearance and shape.

72. Background Subtraction in Highly Dynamic Scenes

Vijay Mahadevan, Nuno Vasconcelos

A new algorithm is proposed for background subtraction in highly dynamic scenes. Background subtraction is equated to the dual problem of saliency detection: background points are those considered not salient by suitable comparison of object and background appearance and dynamics. Drawing inspiration from biological vision, saliency is defined locally, using center-surround computations that measure local feature contrast. A discriminant formulation is adopted, where the saliency of a location is the discriminant power of a set of features with respect to the binary classification problem which opposes center to surround. To account for both motion and appearance, and achieve robustness to highly dynamic backgrounds, these features are spatiotemporal patches, which are modeled as dynamic textures. The resulting background subtraction algorithm is fully unsupervised, requires no training stage to learn background parameters, and depends only on the relative disparity of motion between the center and surround regions. This makes it insensitive to camera motion. The algorithm is tested on challenging video sequences, and shown to outperform various state-of-the-art techniques for background subtraction.

73. A Rank Constrained Continuous Formulation of Multi-frame Multi-target Tracking Problem

Khurram Shafique, Mun Wai Lee, Niels Haering

This paper presents a multi-frame data association algorithm for tracking multiple targets in video sequences. Multi-frame data association involves finding the most probable correspondences between target tracks and measurements (collected over multiple time instances) as well as handling the common tracking problems such as, track initiations and terminations, occlusions, and noisy detections. The problem is known to be NP-Hard for more than two frames. A rank constrained continuous formulation of the problem is presented that can be efficiently solved using nonlinear optimization methods. It is shown that the global and local extrema of the continuous problem respectively coincide with the maximum and the maximal solutions of the discrete counterpart. A scanning window based tracking algorithm is developed using the formulation that performs well under noisy conditions with frequent occlusions and multiple track initiations and terminations. The above claims are supported by experiments and quantitative evaluations using both synthetic and real data under different operating conditions.

74. Fast Algorithms for Large Scale Conditional 3D Prediction

Liefeng Bo, Cristian Sminchisescu, Atul Kanaujia, Dimitris Metaxas

The potential success of discriminative learning approaches to 3D reconstruction relies on the ability to efficiently train predictive algorithms using sufficiently many examples that are representative of the typical configurations encountered in the application domain. Recent research indicates that sparse conditional Bayesian Mixture of Experts (cMoE) models (e.g., BME) are adequate modeling tools that not only provide contextual 3D predictions for problems like human pose reconstruction, but can also represent multiple interpretations that result from depth ambiguities or occlusion. However, training conditional predictors requires sophisticated double-loop algorithms that scale unfavorably with the input dimension and the training set size, thus limiting their usage to 10,000 examples of less, so far. In this paper we present large-scale algorithms, referred to as *fBME*, that combine forward feature selection and bound optimization in order to train probabilistic, BME models, with one order of magnitude more data (100,000 examples and up) and more than one order of magnitude faster. We present several large scale experiments, including monocular evaluation on the HumanEva dataset, demonstrating how the proposed methods overcome the scaling limitations of existing ones.

75. Linear Motion Estimation for Systems of Articulated Planes

Ankur Datta, Yaser Sheikh, Takeo Kanade

In this paper, we describe the explicit application of articulation constraints for estimating the motion of a system of planes. We relate articulations to the relative homography between planes and show that for affine cameras, these articulations translate into linear equality constraints on a linear least squares system, yielding accurate and numerically stable estimates of motion. The global nature of motion estimation allows us to handle areas where there is limited texture information and areas that leave the field of view. Our results demonstrate the accuracy of the algorithm in a variety of cases such as human body tracking, motion estimation of rigid, piecewise planar scenes and motion estimation of triangulated meshes.

10:30am – 12:15pm Oral Session O2A-1: Motion and Tracking (La Perouse)

1. Physical Simulation for Probabilistic Motion Tracking

Marek Vondrak, Leonid Sigal, Odest Chadwicke Jenkins

Human motion tracking is an important problem in computer vision. Most prior approaches have concentrated on efficient inference algorithms and prior motion models; however, few can explicitly account for physical plausibility of recovered motion. The primary purpose of this work is to enforce physical plausibility in the tracking of a single articulated human subject. Towards this end, we propose a fullbody 3D physical simulation-based prior that explicitly incorporates motion control and dynamics into the Bayesian filtering framework. We consider the human's motion to be generated by a “control loop”. In this control loop, Newtonian physics approximates the rigid-body motion dynamics of the human and the environment through the application and integration of forces. Collisions generate interaction forces to prevent physically impossible hypotheses. This allows us to properly model human motion dynamics, ground contact and environment interactions. For efficient inference in the resulting high-dimensional state space, we introduce exemplar-based control strategy to reduce the effective search space. As a result we are able to recover the physically-plausible kinematic and dynamic state of the body from monocular and multi-view imagery. We show, both quantitatively and qualitatively, that our approach performs favorably with respect to standard Bayesian filtering methods.

2. A Mobile Vision System for Robust Multi-Person Tracking

Andreas Ess, Bastian Leibe, Konrad Schindler, Luc Van Gool

We present a mobile vision system for multi-person tracking in busy environments. Specifically, the system integrates continuous visual odometry computation with tracking-by-detection in order to track pedestrians in spite of frequent occlusions and egomotion of the camera rig. To achieve reliable performance under real-world conditions, it has long been advocated to extract and combine as much visual information as possible. We propose a way to closely integrate the vision modules for visual odometry, pedestrian detection, depth estimation, and tracking. The integration naturally leads to several cognitive feedback loops between the modules. Among others, we propose a novel feedback connection from the object detector to visual odometry which utilizes the semantic knowledge of detection to stabilize localization. Feedback loops always carry the danger that erroneous feedback from one module is amplified and causes the entire system to become instable. We therefore incorporate automatic failure detection and recovery, allowing the system to continue when a module becomes unreliable. The approach is experimentally evaluated on several long and difficult video sequences from busy inner-city locations. Our results show that the proposed integration makes it possible to deliver stable tracking performance in scenes of previously infeasible complexity.

3. Motion from Blur

Shengyang Dai, Ying Wu

Motion blur retains some information about motion, based on which motion may be recovered from blurred images. This is a difficult problem, as the situations of motion blur can be quite complicated, such as they may be spacevariant, nonlinear, and local. This paper addresses a very challenging problem: can we recover motion blindly from a single motion-blurred image? A major contribution of this paper is a new finding of an elegant motion blur constraint. Exhibiting a very similar mathematical form as the optical flow constraint, this linear constraint applies locally to pixels in the image. Therefore, a number of challenging problems can be addressed, including estimating global affine motion blur, estimating global rotational motion blur, estimating and segmenting multiple motion blur, and estimating nonparametric motion blur field. Extensive experiments on blur estimation and image deblurring on both synthesized and real data demonstrate the accuracy and general applicability of the proposed approach.

4. People-Tracking-by-Detection and People-Detection-by-Tracking

Mykhaylo Andriluka, Stefan Roth, Bernt Schiele

Both detection and tracking people are challenging problems, especially in complex real world scenes that commonly involve multiple people, complicated occlusions, and cluttered or even moving backgrounds. People detectors have been shown to be able to locate pedestrians even in complex street scenes, but false positives have remained frequent. The identification of particular individuals has remained challenging as well. Tracking methods are able to find a particular individual in image sequences, but are severely challenged by real-world scenarios such as crowded street scenes. In this paper, we combine the advantages of both detection and tracking in a single framework. The approximate articulation of each person is detected in every frame based on local features that model the appearance of individual body parts. Prior knowledge on possible articulations and temporal coherency within a walking cycle are modeled using a hierarchical Gaussian process latent variable model (hGPLVM). We show how the combination of these results improves hypotheses for position and articulation of each person in several subsequent frames. We present experimental results that demonstrate how this allows to detect and track multiple people in cluttered scenes with reoccurring occlusions.

5. Global Data Association for Multi-Object Tracking Using Network Flows

Li Zhang, Yuan Li, Ramakant Nevatia

We propose a network flow based optimization method for data association needed for multiple object tracking. The maximum-a-posteriori (MAP) data association problem is mapped into a cost-flow network with a non-overlap constraint on trajectories. The optimal data association is found by a min-cost flow algorithm in the network. The network is augmented to include an Explicit Occlusion Model (EOM) to track with long-term inter-object occlusions. A solution to the EOM-based network is found by an iterative approach built upon the original algorithm. Initialization and termination of trajectories and potential false observations are modeled by the formulation intrinsically. The method is efficient and does not require hypotheses pruning. Performance is compared with previous results on two public pedestrian datasets to show its improvement.

10:30am – 12:15pm Oral Session O2A-2: Object Detection, Categorization and Recognition (II) (Cook)

1. Epitomic Location Recognition

Kai Ni, Anitha Kannan, Antonio Criminisi, John Winn

This paper presents a novel method for location recognition, which exploits an epitomic representation to achieve both high efficiency and good generalization.

A generative model based on epitomic image analysis captures the appearance and geometric structure of an environment while allowing for variations due to motion, occlusions and non-Lambertian effects. The ability to model translation and scale invariance together with the fusion of diverse visual features yield enhanced generalization with economical training.

Experiments on both existing and new labelled image databases result in recognition accuracy superior to state of the art with real-time computational performance.

2. Beyond Sliding Windows: Object Localization by Efficient Subwindow Search

Christoph H. Lampert, Matthew B. Blaschko, Thomas Hofmann

Most successful object recognition systems rely on binary classification, deciding only if an object is present or not, but not providing information on the actual object location. To perform localization, one can take a sliding window approach, but this strongly increases the computational cost, because the classifier function has to be evaluated over a large set of candidate subwindows.

In this paper, we propose a simple yet powerful branch-and-bound scheme that allows efficient maximization of a large class of classifier functions over all possible subimages. It converges to a globally optimal solution typically in sublinear time. We show how our method is applicable to different object detection and retrieval scenarios. The achieved speedup allows the use of classifiers for localization that formerly were considered too slow for this task, such as SVMs with a spatial pyramid kernel or nearest neighbor classifiers based on the 2 -distance. We demonstrate state-of-the-art performance of the resulting systems on the UIUC Cars dataset, the PASCAL VOC 2006 dataset and in the PASCAL VOC 2007 competition.

3. Closing the Loop in Scene Interpretation

Derek Hoiem, Alexei A. Efros, Martial Hebert

Image understanding involves analyzing many different aspects of the scene. In this paper, we are concerned with how these tasks can be combined in a way that improves the performance of each of them. Inspired by Barrow and Tenenbaum, we present a flexible framework for interfacing scene analysis processes using intrinsic images. Each intrinsic image is a registered map describing one characteristic of the scene. We apply this framework to develop an integrated 3D scene understanding system with estimates of surface orientations, occlusion boundaries, objects, camera viewpoint, and relative depth. Our experiments on a set of 300 outdoor images demonstrate that these tasks reinforce each other, and we illustrate a coherent scene understanding with automatically reconstructed 3D models.

4. A conditional random field for automatic photo editing

Matthew Brand, Patrick Pletscher

We introduce a method for fully automatic touch-up of face images by making inferences about the structure of the scene and undesirable textures in the image. A distribution over image segmentations and labelings is computed via a conditional random field; this distribution controls the application of various local image transforms to regions in the image. Parameters governing both the labeling and transforms are jointly optimized w.r.t. a training set of before-and-after example images. One major advantage of our formulation is the ability to approximately marginalize over all possible labelings and thus exploit much or most of the information in the distribution; this yields better results than MAP inference. We demonstrate with a system that is trained to correct red-eye, reduce specularities, and remove acne and other blemishes from faces, showing results with test images scavenged from acne-themed internet message boards.

5. Transductive Object Cutout

Jingyu Cui, Qiong Yang, Fang Wen, Qiying Wu, Changshui Zhang, Xiaoou Tang

In this paper, we address the issue of transducing the object cutout model from an example image to novel image instances. We observe that although object and background are very likely to contain similar colors in natural images, it is much less probable that they share similar color configurations. Motivated by this observation, we propose a local color pattern model to characterize the color configuration in a robust way. Additionally, we propose an edge profile model to modulate the contrast of the image, which enhances edges along object boundaries and attenuates edges inside object or background. The local color pattern model and edge model are integrated in a graph-cut framework. Higher accuracy and improved robustness of the proposed method are demonstrated through experimental comparison with state-of-the-art algorithms.

1:45pm – 3:45pm Poster Session P2P-1: Object Recognition and Color & Texture (Summit Hall)

1. Simultaneous Learning of a Discriminative Projection and Prototypes for Nearest-Neighbor Classification

Mauricio Villegas, Roberto Paredes

Computer vision and image recognition research have a great interest in dimensionality reduction techniques. Generally these techniques are independent of the classifier being used and the learning of the classifier is carried out after the dimensionality reduction is performed, possibly discarding valuable information. In this paper we propose an iterative algorithm that simultaneously learns a linear projection base and a reduced set of prototypes optimized for the Nearest-Neighbor classifier. The algorithm is derived by minimizing a suitable estimation of the classification error probability. The proposed approach is assessed through a series of experiments showing a good behavior and a real potential for practical applications.

2. Margin-Based Discriminant Dimensionality Reduction for Visual Recognition

Hakan Cevikalp, Bill Triggs, Frédéric Jurie, Robi Polikar

Nearest neighbour classifiers and related kernel methods often perform poorly in high dimensional problems because it is infeasible to include enough training samples to cover the class regions densely. In such cases, test samples often fall into gaps between training samples where the nearest neighbours are too distant to be good indicators of class membership. One solution is to project the data onto a discriminative lower dimensional subspace. We propose a gap-resistant nonparametric method for finding such subspaces: first the gaps are filled by building a convex model of the region spanned by each class -- we test the affine and convex hulls and the bounding disk of the class training samples -- then a set of highly discriminative directions is found by building and decomposing a scatter matrix of weighted displacement vectors from training examples to nearby rival class regions. The weights are chosen to focus attention on narrow margin cases while still allowing more diversity and hence more discriminability than the 1D linear Support Vector Machine (SVM) projection. Experimental results on several face and object recognition datasets show that the method finds effective projections, allowing simple classifiers such as nearest neighbours to work well in the low dimensional reduced space.

3. A Mixed Generative-Discriminative Framework for Pedestrian Classification

Markus Enzweiler, Dariu M. Gavrila

This paper presents a novel approach to pedestrian classification which involves utilizing the synthesized virtual samples of a learned generative model to enhance the classification performance of a discriminative model. Our generative model captures prior knowledge about the pedestrian class in terms of a number of probabilistic shape and texture models, each attuned to a particular pedestrian pose. Active learning provides the link between the generative and discriminative model, in the sense that the former is selectively sampled such that the training process is guided towards the most informative samples of the latter.

In large-scale experiments on real-world datasets of tens of thousands of samples, we demonstrate a significant improvement in classification performance of the combined generative-discriminative approach over the discriminative-only approach (the latter exemplified by a neural network with local receptive fields and a support vector machine using Haar wavelet features).

4. Learning Coupled Conditional Random Field for Image Decomposition with Application on Object Categorization

Xiaoxu Ma, W. Eric L. Grimson

This paper proposes a computational system of object categorization based on decomposition and adaptive fusion of visual information. A coupled Conditional Random Field is developed to model the interaction between low level cues of contour and texture, and to decompose contour and texture in natural images. The advantages of using coupled rather than single-layer Random Fields are demonstrated with model learning and evaluation. Multiple decomposed visual cues are adaptively combined for object categorization to fully leverage different discriminative cues for different classes. Experimental results show that the proposed computational model of "recognition-through-decomposition-and-fusion" achieves better performance than most of the state-of-the-art methods, especially when only a limited number of training samples are available.

5. Regularizing 3D Medial Axis Using Medial Scaffold Transforms

Ming-Ching Chang, Benjamin Kimia

This paper addresses a key bottleneck in the use of the 3D medial axis (MA) representation, namely, how the complex MA structure can be regularized so that similar, within-category 3D shapes yield similar 3D MA that are distinct from the non-category shapes. We rely on previous work which (i) constructs a hierarchical MA hypergraph, the medial scaffold (MS), and (ii) the theoretical classification of the instabilities of this structure, or transitions (sudden topological changes due to a small perturbation). The shapes at transition point are degenerate. Our approach is to recognize the transitions which are close-by to a given shape and transform the shape to this transition point, and repeat until no close-by transitions exists. This move towards degeneracy is the basis of simplification of shape. We derive 11 transforms from 7 transitions and follow a greedy scheme in applying the transform. The results show that the simplified MA preserves with-in-category similarity, thus indicating its potential use in various applications including shape analysis, manipulation, and matching.

6. From Appearance to Context-Based Recognition: Dense Labeling in Small Images

Devi Parikh, C. Lawrence Zitnick, Tsuhan Chen

Traditionally, object recognition is performed based solely on the appearance of the object. However, relevant information also exists in the scene surrounding the object. As supported by our human studies, this contextual information is necessary for accurate recognition in low resolution images. This scenario with impoverished appearance information, as opposed to using images of higher resolution, provides an appropriate venue for studying the role of context in recognition. In this paper, we explore the role of context for dense scene labeling in small images. Given a segmentation of an image, our algorithm assigns each segment to an object category based on the segment's appearance and contextual information. We explicitly model context between object categories through the use of relative location and relative scale, in addition to co-occurrence. We perform recognition tests on low and high resolution images, which vary significantly in the amount of appearance information present, using just the object appearance information, the combination of appearance and context, as well as just context without object appearance information (blind recognition). We also perform these tests in human studies and analyze our findings to reveal interesting patterns. With the use of our context model, our algorithm achieves state-of-the-art performance on MSRC and Corel datasets.

7. Exploiting Side Information in Locality Preserving Projection

Senjian An, Wanquan Liu, Svetha Venkatesh

Even if the class label information is unknown, side information represents some equivalence constraints between pairs of patterns, indicating whether pairs originate from the same class. Exploiting side information, we develop algorithms to preserve both the **intra-class and inter-class local structures**. This new type of locality preserving projection (LPP), called LPP with side information (LPPSI), preserves the data's local structure in the sense that the close, similar training patterns will be kept close, whilst the close but dissimilar ones are separated. Our algorithms balance these conflicting requirements, and we further improve this technique using kernel methods. Experiments conducted on popular face databases demonstrate that the proposed algorithm significantly outperforms LPP. Further, we show that the performance of our algorithm with partial side information (that is, using only small amount of pair-wise similarity/dissimilarity information during training) is comparable with that when using full side information. We conclude that exploiting side information by preserving both similar and dissimilar local structures of the data significantly improves performance.

8. A Discriminatively Trained, Multiscale, Deformable Part Model

Pedro Felzenszwalb, David McAllester, Deva Ramanan

This paper describes a discriminatively trained, multiscale, deformable part model for object detection. Our system achieves a two-fold improvement in average precision over the best performance in the 2006 PASCAL person detection challenge. It also outperforms the best results in the 2007 challenge in ten out of twenty categories. The system relies heavily on deformable parts. While deformable part models have become quite popular, their value had not been demonstrated on difficult benchmarks such as the PASCAL challenge. Our system also relies heavily on new methods for discriminative training. We combine a margin-sensitive approach for data mining hard negative examples with a formalism we call latent SVM. A latent SVM, like a hidden CRF, leads to a non-convex training problem. However, a latent SVM is semi-convex and the training problem becomes convex once latent information is specified for the positive examples. We believe that our training methods will eventually make possible the effective use of more latent information such as hierarchical (grammar) models and models involving latent three dimensional pose.

9. In Defense of Nearest-Neighbor Based Image Classification

Oren Boiman, Eli Shechtman, Michal Irani

State-of-the-art image classification methods require an intensive learning/training stage (using SVM, Boosting, etc.) In contrast, non-parametric Nearest-Neighbor (NN) based image classifiers require no training time and have other favorable properties. However, the large performance gap between these two families of approaches rendered NN-based image classifiers useless.

We claim that the effectiveness of non-parametric NN-based image classification has been considerably under-valued. We argue that two practices commonly used in image classification methods, have led to the inferior performance of NN-based image classifiers: (i) Quantization of local image descriptors (used to generate “bags-of-words”, codebooks). (ii) Computation of `Image-to-Image' distance, instead of `Image-to-Class' distance.

We propose a trivial NN-based classifier -- NBNN, (Naive-Bayes Nearest-Neighbor), which employs NN-distances in the space of the local image descriptors (and not in the space of images). NBNN computes direct `Image-to-Class' distances without descriptor quantization. We further show that under the Naive-Bayes assumption, the theoretically optimal image classifier can be accurately approximated by NBNN.

Although NBNN is extremely simple, efficient, and requires no learning/training phase, its performance ranks among the top leading learning-based image classifiers. Empirical comparisons are shown on several challenging databases (Caltech-101, Caltech-256, and Graz-01).

10. Enhanced Biologically Inspired Model

Yongzhen Huang, kaiqi Huang, Dacheng Tao, Liangsheng Wang, Tieniu Tan, Xuelong Li

It has been demonstrated by Serre et al. that the biologically inspired model (BIM) is effective for object recognition. It outperforms many state-of-the-art methods in challenging databases. However, BIM has the following three problems: a very heavy computational cost due to dense input, a disputable pooling operation in modeling relations of the visual cortex, and blind feature selection in a feedforward framework. To solve these problems, we develop an enhanced BIM (EBIM), which removes uninformative input by imposing sparsity constraints, utilizes a novel local weighted pooling operation with stronger physiological motivations, and applies a feedback procedure that selects effective features for combination. Empirical studies on the CalTech5 database and CalTech101 database show that EBIM is more effective and efficient than BIM. We also apply EBIM to the MIT-CBCL street scene database to show it achieves comparable performance in comparison with the current best performance. Moreover, the new system can process images with resolution 128 × 128 at a rate of 50 frames per second and enhances the speed 20 times at least in comparison with BIM in common applications.

11. A similarity measure between unordered vector sets with application to image categorization

Yan Liu, Florent Perronnin

We present a novel approach to compute the similarity between two unordered variable-sized vector sets. To solve this problem, several authors have proposed to model each vector set with a Gaussian mixture model (GMM) and to compute a probabilistic measure of similarity between the GMMs. The main contribution of this paper is to model each vector set with a GMM adapted from a common “universal” GMM using the maximum a posteriori (MAP) criterion. The advantages of this approach are twofold. MAP provides a more accurate estimate of the GMM parameters compared to standard maximum likelihood estimation (MLE) in the challenging case where the cardinality of the vector set is small. Moreover, there is a correspondence between the Gaussians of two GMMs adapted from a common distribution and one can take advantage of this fact to compute efficiently the probabilistic similarity. This work is applied to the image categorization problem: images are modeled as bags of low-level features and classification is performed using a kernel classifier based on the proposed similarity measure. Experimental results on the PASCAL VOC 2006 and VOC 2007 databases show the excellent performance of our approach.

12. Loose Shape Model for Discriminative Learning of Object Categories

Margarita Osadchy, Morash Elran

We consider the problem of visual categorization with minimal supervision during training. We propose a part-based model that loosely captures structural information. We represent images as a collection of parts characterized by an appearance codeword from a visual vocabulary and by a neighborhood context, organized in an ordered set of bag-of-features representations. These bags are computed in a local overlapping areas around the part. A semantic distance between images is obtained by matching parts associated with the same codeword using their context distributions. The classification is done using SVM with the kernel obtained from the proposed distance. The experiments show that our method outperforms all the classification methods from the PASCAL challenge on half of the VOC2006 categories and has the best average EER. It also outperforms the constellation model learned via boosting, as proposed by Bar-Hillel et al. on their data set, which contains more rigid objects.

13. Scene Understanding with Discriminative Structured Prediction

Jinhui Yuan, Jianmin Li, Bo Zhang

Spatial priors play crucial roles in many high-level vision tasks, e.g., scene understanding. Usually, learning spatial priors relies on training a structured output model. In this paper, two special cases of discriminative structured output model, i.e., Conditional Random Fields (CRFs) and Max-margin Markov Networks (M^3N), are demonstrated to perform image scene understanding. The two models are empirically compared in a fair manner, i.e., using the common feature representation and the same optimization algorithm. Particularly, we adopt online Exponentiated Gradient (EG) algorithm to solve the convex duals of both models. We describe the general procedure of EG algorithm and present a two-stage training procedure to overcome the degeneration of EG when exact inference is intractable. Experiments on a large scale image region annotation task are carried out. The results show that both models yield encouraging results but CRFs slightly outperforms M^3N.

14. Automatic Face Naming with Caption-based Supervision

Matthieu Guillaumin, Thomas Mensink, Jakob Verbeek, Cordelia Schmid

We consider two scenarios of naming people in databases of news photos with captions: (i) finding faces of a single person, and (ii) assigning names to all faces. We combine an initial text-based step, that restricts the name assigned to a face to the set of names appearing in the caption, with a second step that analyzes visual features of faces. By searching for groups of highly similar faces that can be associated with a name, the results of purely text-based search can be greatly ameliorated. We improve a recent graph-based approach, in which nodes correspond to faces and edges connect highly similar faces. We introduce constraints when optimizing the objective function, and propose improvements in the low-level methods used to construct the graphs. Furthermore, we generalize the graph-based approach to face naming in the full data set. In this multi-person naming case the optimization quickly becomes computationally demanding, and we present an important speed-up using graph-flows to compute the optimal name assignments in documents. Generative models have previously been proposed to solve the multi-person naming task. We compare the generative and graph-based methods in both scenarios, and find significantly better performance using the graph-based methods in both cases.

15. Learning-based Face Hallucination in DCT Domain

Wei Zhang, Wai-Kuen Cham

In this paper, we propose a novel learning-based face hallucination framework built in DCT domain, which can recover the high-resolution face image from a single low-resolution one. Unlike most previous learning-based work, our approach addresses the face hallucination problem from a different angle. In details, the problem is formulated as inferring DCT coefficients in frequency domain instead of estimating pixel intensities in spatial domain. Experimental results show that DC coefficients can be estimated fairly accurately by simple interpolation-based methods. AC coefficients, which contain the information of local features of face image, cannot be estimated well using interpolation. We propose a method to infer AC coefficients by introducing an efficient learning-based inference model. Moreover, the proposed framework can lead to significant savings in memory and computation cost since the redundancy of the training set is reduced a lot by clustering. Experimental results demonstrate that our approach is very effective to produce hallucinated face images with high quality.

16. Automatic symmetry plane estimation of bilateral objects in point clouds

Benoît Combès, Robin Hennessy, John Waddington, Neil Roberts, Sylvain Prima

In this paper, the problem of estimating automatically the symmetry plane of bilateral objects (having perfect or imperfect mirror symmetry) in point clouds is reexamined. Classical methods, mostly based on the ICP algorithm, are shown to be limited and complicated by an inappropriate parameterization of the problem. First, we show how an adequate parameterization, used in an ICP-like scheme, can lead to a simpler, more accurate and faster algorithm. Then, using this parameterization, we reinterpret the problem in a probabilistic framework, and use the maximum likelihood principle to define the optimal symmetry plane. This problem can be solved efficiently using an EM algorithm. The resulting iterative scheme can be seen as an ICP-like algorithm with multiple matches between the two sides of the object. This new algorithm, implemented using a multiscale, multiresolution approach, is evaluated in terms of accuracy, robustness and speed on ground truth data, and some results on real data are presented.

17. Correspondences between Parts of Shapes with Particle Filters

Rolf Lakaemper, Marc Sobel

Given two shapes, the correspondence between distinct visual features is the basis for most alignment processes and shape similarity measures. This paper presents an approach introducing particle filters to establish perceptually correct correspondences between point sets representing shapes. Local shape feature descriptors are used to establish correspondence probabilities. The global correspondence structure is calculated using additional constraints based on domain knowledge. Domain knowledge is characterized as prior distributions expressing hypotheses about the global relationships between shapes. These hypotheses are generated during the iterative particle filtering process. Experiments using standard alignment techniques, based on the given correspondence relationships, demonstrate the advantages of this approach.

18. Context-Dependent Kernel Design for Object Matching and Recognition

Hichem Sahbi, Jean-Yves Audibert, Jaonary Rabarisoa, Renaud Keriven

The success of kernel methods including support vector networks (SVMs) strongly depends on the design of appropriate kernels. While initially kernels were designed in order to handle fixed-length data, their extension to unordered, variable-length data became more than necessary for real pattern recognition problems such as object recognition and bioinformatics.

We focus in this paper on object recognition using a new type of kernel referred to as “context-dependent”. Objects, seen as constellations of local features (interest points, regions, etc.), are matched by minimizing an energy function mixing (1) a fidelity term which measures the quality of feature matching, (2) a neighborhood criteria which captures the object geometry and (3) a regularization term. We will show that the fixed-point of this energy is a “context-dependent” kernel (“CDK”) which also satisfies the Mercer condition. Experiments conducted on object recognition show that when plugging our kernel in SVMs, we clearly outperform SVMs with “context-free” kernels.

19. Matching Vehicles under Large Pose Transformations using Approximate 3D Models and Piecewise MRF Model

Yanlin Guo, Cen Rao, Janet Kim, Harpreet Sawhney, Rakesh Kumar, Supun Samarasekera

We propose a robust object recognition method based on approximate 3D models that can effectively match objects under large viewpoint changes and partial occlusion. The specific problem we solve is: given two views of an object, determine if the views are for the same or different object. Our domain of interest is vehicles, but the approach can be generalized to other man-made rigid objects. A key contribution of our approach is the use of approximate models with locally and globally constrained rendering to determine matching objects. We utilize a compact set of 3D models to provide geometry constraints and transfer appearance features for object matching across disparate viewpoints. The closest model from the set, together with its poses with respect to the data, is used to render an object both at pixel (local) level and region/part (global) level. Especially, symmetry and semantic part ownership are used to extrapolate appearance information. A piecewise Markov Random Field (MRF) model is employed to combine observations obtained from local pixel and global region level. Belief Propagation (BP) with reduced memory requirement is employed to solve the MRF model effectively. No training is required, and a realistic object image in a disparate viewpoint can be obtained from as few as just one image. Experimental results on vehicle data from multiple sensor platforms demonstrate the efficacy of our method.

20. Estimating Age, Gender, and Identity using First Name Priors

Andrew Gallagher, Tsuhan Chen

Recognizing people in images is one of the foremost challenges in computer vision. It is important to remember that consumer photography has a highly social aspect. The photographer captures images not in a random fashion, but rather to remember or document meaningful events in her life. The culture of the society of which the photographer is a part provides a strong context for recognizing the content of the captured images.

We demonstrate one aspect of this cultural context by recognizing people from first names. The distribution of first names chosen for newborn babies evolves with time and is gender-specific. As a result, a first name provides a strong prior for describing the individual. Specifically, we use the U.S. Social Security Administration baby name database to learn priors for gender and age for 6693 first names.

Most face recognition methods do not even consider the name of the individual of interest, or the name is treated merely as an identifier that provides no information about appearance. In contrast, we combine image-based gender and age classifiers with the cultural context information provided by first names to recognize people with no labeled examples. Our model uses image-based age and gender estimates for assigning first names to people and in turn, the age and gender estimates are improved.

21. Local Tensor Descriptor from Micro-deformation Analysis

Bangsheng Cheng

This paper proposes a novel method called micro-deformation analysis to analyze and describe local image structures. This method is a general analytic tool and can be applied to any high-dimensional scalar or vector functions. We derive the tensor matrix from this method as the descriptor to represent the information within local image patches. Our experimental results suggest that we can design low-dimensional local tensor descriptors with performance comparable to the popular SIFT descriptor, which is the state-of-the-art feature descriptor used for object recognition and categorization.

22. Visual Synset: Towards a Higher-level Visual Representation

Yan-Tao Zheng, Ming Zhao, Shi-Yong Neo, Tat-Seng Chua, Qi Tian

We present a higher-level visual representation, visual synset, for object categorization. The visual synset improves the traditional bag of words representation with better discrimination and invariance power. First, the approach strengthens the inter-class discrimination power by constructing an intermediate visual descriptor, delta visual phrase, from frequently co-occurring visual word-set with similar spatial context. Second, the approach achieves better intra-class invariance power, by clustering delta visual phrases into visual synset, based their probabilistic 'semantics', i.e., class probability distribution. Hence, the resulting visual synset can partially bridge the visual differences of images of same class. The tests on Caltech-101 and Pascal-VOC 05 dataset demonstrated that the proposed image representation can achieve good accuracies.

23. ANSIG - An Analytic Signature for Permutation-Invariant Two-Dimensional Shape Representation

José J. Rodrigues, Pedro M. Q. Aguiar, João M. F. Xavier

Many applications require a computer representation of 2D shape, usually described by a set of 2D points. The challenge of this representation is that it must not only capture the characteristics of the shape but also be invariant to relevant transformations. Invariance to geometric transformations, such as translation, rotation and scale, has received attention in the past, usually under the assumption that the points are previously labeled, i.e., that the shape is characterized by an ordered set of landmarks. However, in many practical scenarios the landmarks are obtained from an automatic process, e.g., edge/corner detection, thus without natural ordering. In this paper, we represent 2D shapes in a way that is invariant to the permutation of the landmarks. Within our framework, a shape is mapped to an analytic function on the complex plane, leading to what we call its analytic signature (ANSIG). We show that different shapes lead to different ANSIGs but that shapes that differ by a permutation of the landmarks lead to the same ANSIG, i.e., that our representation is a maximal invariant with respect to the permutation group. To store an ANSIG, it suffices to sample it along a closed contour in the complex plane. We further show how easy it is to factor out geometric transformations when comparing shapes using the ANSIG representation. We illustrate the ANSIG capabilities in shape-based image classification.

24. Adaptive and Compact Shape Descriptor by Progressive Feature Combination and Selection with Boosting

Cheng Chen, Yueting Zhuang, Jun Xiao, Fei Wu

Many types of shape descriptors have been proposed for 2D shape analysis, but most of them consist of component features that are not adapted to specific problems. This has two drawbacks. First, computation is wasted on the irrelevant components; second, the accuracy is impaired. This paper proposes an effective method that generates compact descriptors adapted to specific problems in hand, where each component of the new descriptor is a linear combination of the components in some classic descriptors. A progressive strategy is used to construct and select the most suitable linear combinations in successive rounds, where a variant of Adaboost is employed to ensure the optimum of the selected combinations in each round. Experiments show that our method effectively generates adaptive and compact descriptors for typical applications such as shape classification and retrieval.

25. Viewpoint-Independent Object Class Detection using 3D Feature Maps

Joerg Liebelt, Cordelia Schmid, Klaus Schertler

This paper presents a 3D approach to multi-view object class detection. Most existing approaches recognize object classes for a particular viewpoint or combine classifiers for a few discrete views. We propose instead to build 3D representations of object classes which allow to handle viewpoint changes and intra-class variability. Our approach extracts a set of pose and class discriminant features from synthetic 3D object models using a filtering procedure, evaluates their suitability for matching to real image data and represents them by their appearance and 3D position. We term these representations 3D Feature Maps. For recognizing an object class in an image we match the synthetic descriptors to the real ones in a 3D voting scheme. Geometric coherence is reinforced by means of a robust pose estimation which yields a 3D bounding box in addition to the 2D localization. The precision of the 3D pose estimation is evaluated on a set of images of a calibrated scene. The 2D localization is evaluated on the PASCAL 2006 dataset for motorbikes and cars, showing that its performance can compete with state-of-the-art 2D object detectors.

26. Unsupervised Feature Selection via Distributed Coding for Multi-view Object Recognition

C. Mario Christoudias, Raquel Urtasun, Trevor Darrell

Object recognition accuracy can be improved when information from multiple views is integrated, but information in each view can often be highly redundant. We consider the problem of distributed object recognition or indexing from multiple cameras, where the computational power available at each camera sensor is limited and communication between cameras is prohibitively expensive. In this scenario, it is desirable to avoid sending redundant visual features from multiple views. Traditional supervised feature selection approaches are inapplicable as the class label is unknown at each camera. In this paper we propose an unsupervised multi-view feature selection algorithm based on a distributed coding approach. With our method, a Gaussian Process model of the joint view statistics is used at the receiver to obtain a joint encoding of the views without directly sharing information across encoders. We demonstrate our approach on recognition and indexing tasks with multi-view image databases and show that our method compares favorably to an independent encoding of the features from each camera.

27. Dynamic Visual Category Learning

Tom Yeh, Trevor Darrell

Dynamic visual category learning calls for efficient adaptation as new training images become available or new categories are defined, existing training images or categories become modified or obsolete, or when categories are divided into subcategories or merged together. We develop novel methods for efficient incremental learning of SVM-based visual category classifiers to handle such dynamic tasks. Our method exploits previous classifier estimates to more efficiently learn the optimal parameters for the current set of training images and categories. We show empirically that for dynamic visual category tasks, our incremental learning methods are significantly faster than batch retraining.

28. Randomized Trees for Human Pose Detection

Grégory Rogez, Jonathan Rihan, Srikumar Ramalingam, Carlos Orrite, Philip H. S. Torr

This paper addresses human pose recognition from video sequences by formulating it as a classification problem. Unlike much previous work we do not make any assumptions on the availability of clean segmentation. The first step of this work consists in a novel method of aligning the training images using 3D Mocap data. Next we define classes by discretizing a 2D manifold whose two dimensions are camera viewpoint and actions. Our main contribution is a pose detection algorithm based on random forests. A bottom-up approach is followed to build a decision tree by recursively clustering and merging the classes at each level. For each node of the decision tree we build a list of potentially discriminative features using the alignment of training images; in this paper we consider Histograms of Orientated Gradient (HOG). We finally grow an ensemble of trees by randomly sampling one of the selected HOG blocks at each node. Our proposed approach gives promising results with both fixed and moving cameras.

29. Combining Brain Computer Interfaces with Vision for Object Categorization

Ashish Kapoor, Pradeep Shenoy, Desney Tan

Human-aided computing proposes using information measured directly from the human brain in order to perform useful tasks. In this paper, we extend this idea by fusing computer vision-based processing and processing done by the human brain in order to build more effective object categorization systems. Specifically, we use an electroencephalograph (EEG) device to measure the subconscious cognitive processing that occurs in the brain as users see images, even when they are not trying to explicitly classify them. We present a novel framework that combines a discriminative visual category recognition system based on the Pyramid Match Kernel (PMK) with information derived from EEG measurements as users view images. We propose a fast convex kernel alignment algorithm to effectively combine the two sources of information. Our approach is validated with experiments using real-world data, where we show significant gains in classification accuracy. We analyze the properties of this information fusion method by examining the relative contributions of the two modalities, the errors arising from each source, and the stability of the combination in repeated experiments.

30. Relaxed Matching Kernels for Object Recognition

Andrea Vedaldi, Stefano Soatto

The popular bag-of-features representation for object recognition collects signatures of local image patches and discards spatial information. Some have recently attempted to at least partially overcome this limitation, for instance by “spatial pyramids” and “proximity” kernels. We introduce the general formalism of “relaxed matching kernels” (RMKs) that includes such approaches as special cases, allow us to derive useful general properties of these kernels, and to introduce new ones. As an example, we introduce a kernel based on matching graphs of features and one based on matching information-compressed features. We show that all RMKs are competitive and outperform in several cases recently published state-of-the-art results on standard datasets. However, we also show that a proper implementation of a baseline bag-of-features algorithm can be extremely competitive, and outperform the other methods in some cases.

31. Unsupervised Learning of Visual Taxonomies

Evgeniy Bart, Ian Porteous, Pietro Perona, Max Welling

As more images and categories become available, organizing them becomes crucial. We present a novel statistical method for organizing a collection of images into a tree-shaped hierarchy. The method employs a non-parametric Bayesian model and is completely unsupervised. Each image is associated with a path through a tree. Similar images share initial segments of their paths and therefore have a smaller distance from each other. Each internal node in the hierarchy represents information that is common to images whose paths pass through that node, thus providing a compact image representation. Our experiments show that a disorganized collection of images will be organized into an intuitive taxonomy. Furthermore, we find that the taxonomy allows good image categorization and, in this respect, is superior to the popular LDA model.

32. Filtering Internet Image Search Results Towards Keyword Based Category Recognition

Kamil Wnuk, Stefano Soatto

In this work we aim to capitalize on the availability of Internet image search engines to automatically create image training sets from user provided queries. This problem is particularly difficult due to the low precision of image search results. Unlike many existing dataset gathering approaches, we do not assume a category model based on a small subset of the noisy data or an ad-hoc validation set. Instead we use a nonparametric measure of strangeness in the space of holistic image representations, and perform an iterative feature elimination algorithm to remove the most strange examples from the category. This is the equivalent of keeping only features that are found to be consistent with others in the class. We show that applying our method to image search data before training improves average recognition performance, and demonstrate that we obtain comparative precision and recall results to the current state of the art, all the while maintaining a significantly simpler approach. In the process we also extend the strangeness-based feature elimination algorithm to automatically select good threshold values and perform filtering of a single class when the background is given.

33. Unsupervised Discovery of Visual Object Class Hierarchies

Josef Sivic, Bryan Russell, Andrew Zisserman, William T. Freeman, Alexei A. Efros

Objects in the world can be arranged into a hierarchy based on their semantic meaning (e.g., organism -- animal -- feline -- cat). What about defining a hierarchy based on the *visual appearance* of objects? This paper investigates ways to automatically discover a hierarchical structure for the visual world from a collection of unlabeled images. Previous approaches for unsupervised object and scene discovery focused on partitioning the visual data into a set of non-overlapping classes of equal granularity. In this work, we propose to group visual objects using a multi-layer hierarchy tree that is based on common visual elements. This is achieved by adapting to the visual domain the generative Hierarchical Latent Dirichlet Allocation (hLDA) model previously used for unsupervised discovery of topic hierarchies in text. Images are modeled using quantized local image regions as analogues to words in text. Employing the multiple segmentation framework of Russell et al., we show that meaningful object hierarchies, together with object segmentations, can be automatically learned from unlabeled and unsegmented image collections without supervision. We demonstrate improved object classification and localization performance using hLDA over the previous non-hierarchical method on the MSRC dataset.

34. Improving Local Learning for Object Categorization by Exploring the Effects of Ranking

Tien-Lung Chang, Tyng-Luh Liu, Jen-Hui Chuang

Local learning for classification is useful in dealing with various vision problems. One key factor for such approaches to be effective is to find good neighbors for the learning procedure. In this work, we describe a novel method to rank neighbors by learning a local distance function, and meanwhile to derive the local distance function by focusing on the high-ranked neighbors. The two aspects of considerations can be elegantly coupled through a well-defined objective function, motivated by a supervised ranking method called P-Norm Push. While the local distance functions are learned independently, they can be reshaped altogether so that their values can be directly compared. We apply the proposed method to the Caltech-101 dataset, and demonstrate the use of proper neighbors can improve the performance of classification techniques based on nearest-neighbor selection.

35. An Experimental Study of Employing Visual Appearance as a Phenotype

Lior Wolf, Yoni Donner

Visual and non-visual data are often related through complex, indirect links, thus making the prediction of one from the other difficult. Examples include the partially-understood connections between firing of V1 neurons and visual stimuli, the coupling between recorded speech and video of the corresponding lip movements, and the attempts to infer criminal intentions from surveillance videos.

In this study, we explore the exploitation of the visual/non-visual relation between genetic sequences and visual appearance. This exploitation is currently considered infeasible due to the many hidden variables and unknown factors involved, the considerable variability and noise that exist in images and the high-dimensionality of the data.

Despite the difficulties, we show convincing evidence that the application of correlations between genotype and visual phenotype for identification is feasible with current technologies. To this end, we employ sensitive forced-matching tests, that can accurately detect correlations between data sets. These tests are used to compare the performance of several existing algorithms, as well as novel ones that we have designed for the task.

36. Manifold Learning using Robust Graph Laplacian for Interactive Image Search

Hichem Sahbi, Patrick Etyngier, Jean-Yves Audibert, Renaud Keriven

Interactive image search or relevance feedback is the process which helps a user refining his query and finding difficult target categories. This consists in partially labeling a very small fraction of an image database and iteratively refining a decision rule using both the labeled and unlabeled data. Training of this decision rule is referred to as transductive learning.

Our work is an original approach for relevance feedback based on Graph Laplacian. We introduce a new Graph Laplacian which makes it possible to robustly learn the embedding, of the manifold enclosing the dataset, via a diffusion map. Our approach is three-folds: it allows us (i) to integrate all the unlabeled images in the decision process (ii) to robustly capture the topology of the image set and (iii) to perform the search process inside the manifold. Relevance feedback experiments were conducted on simple databases including Olivetti and Swedish as well as challenging and large scale databases including Corel. Comparisons show clear and consistent gain, of our graph Laplacian method, with respect to state-of-the art relevance feedback approaches.

37. Connected Segmentation Tree -- A Joint Representation of Region Layout and Hierarchy

Narendra Ahuja, Todorovic Sinisa

This paper proposes a new object representation, called Connected Segmentation Tree (CST), which captures canonical characteristics of the object in terms of the photometric, geometric, and spatial adjacency and containment properties of its constituent image regions. CST is obtained by augmenting the object's segmentation tree (ST) with inter-region neighbor links, in addition to their recursive embedding structure already present in ST. This makes CST a hierarchy of region adjacency graphs. A region's neighbors are computed using an extension to regions of the Voronoi diagram for point patterns. Unsupervised learning of the CST model of a category is formulated as matching the CST graph representations of unlabeled training images, and fusing their maximally matching subgraphs. A new learning algorithm is proposed that optimizes the model structure by *simultaneously* searching for both the most salient nodes (regions) and the most salient edges (containment and neighbor relationships of regions) across the image graphs. Matching of the category model to the CST of a new image results in simultaneous detection, segmentation and recognition of all occurrences of the category, and a semantic explanation of these results.

38. Where am I: Place Instance and Category Recognition using Spatial PACT

Jianxin Wu, James Rehg

We introduce spatial PACT (Principal component Analysis of Census Transform histograms), a new representation for recognizing instances and categories of places or scenes. Both place instance recognition (“I am in Room 113”) and category recognition (“I am in an office”) have been widely researched. Features that have different discriminative power/invariance tradeoff have been used separately for the two tasks. PACT captures local structures of an image through the Census Transform (CT), while large-scale structures are captured by the strong correlation between neighboring CT values and the histogram. The PCA operation ignores noise in the histogram distribution, computes important “primitive shapes”, and results in a compact representation. Spatial PACT, a spatial pyramid of PACT, further incorporates global structures in the image. Our experiments demonstrate that spatial PACT outperforms the current state-of-the-art in several place and scene recognition, and shape matching datasets. Besides, spatial PACT is easy to implement. It has nearly no parameter to tune, and evaluates extremely fast.

39. Action Recognition with Motion-Appearance Vocabulary Forest

Krystian Mikolajczyk, Uemura Hirofumi

In this paper we propose an approach for action recognition based on a vocabulary forest of local motion-appearance features. Large numbers of features with associated motion vectors are extracted from action data and are represented by many vocabulary trees. Features from a query sequence are matched to the trees and vote for action categories and their locations. Large number of trees make the process efficient and robust. The system is capable of simultaneous categorization and localization of actions using only a few frames per sequence. The approach obtains excellent performance on standard action recognition sequences. We perform large scale experiments on 17 challenging real action categories from olympic games. We demonstrate the robustness of our method to appearance variations, camera motion, scale change, asymmetric actions, background clutter and occlusion.

40. Semi-Supervised Boosting using Visual Similarity Learning

Leistner Christian, Grabner Helmut, Bischof Horst

The required amount of labeled training data for object detection and classification is a major drawback of current methods. Combining labeled and unlabeled data via semi-supervised learning holds the promise to ease the tedious and time consuming labeling effort. This paper presents a novel semi-supervised learning method which combines the power of learned similarity functions and classifiers. The approach capable of exploiting both labeled and unlabeled data is formulated in a boosting framework. One classifier (the learned similarity) serves as a prior which is steadily improved via training a second classifier on labeled and unlabeled samples. We demonstrate the approach on challenging computer vision applications. First, we show how we can train a classifier using only a few labeled samples and many unlabeled data. Second, we improve (specialize) a state-of-the-art detector by using labeled and unlabeled data.

41. Classification using Intersection Kernel Support Vector Machines is Efficient

Subhransu Maji, Alexander Berg, Jitendra Malik

Straightforward classification using kernelized SVMs requires evaluating the kernel for a test vector and each of the support vectors. For a class of kernels we show that one can do this much more efficiently. In particular we show that one can build histogram intersection kernel SVMs (IKSVMs) with runtime complexity of the classifier logarithmic in the number of support vectors as opposed to linear for the standard approach. We further show that by precomputing auxiliary tables we can construct an approximate classifier with constant runtime and space requirements, independent of the number of support vectors, with negligible loss in classification accuracy on various tasks. This approximation also applies to $1 - \chi^2$ and other kernels of similar form.

We also introduce novel features based on a multi-level histograms of oriented edge energy and present experiments on various detection datasets. On the INRIA pedestrian dataset an approximate IKSVM classifier based on these features has the current best performance, with a miss rate 13% lower at 10^{-6} False Positive Per Window than the linear SVM detector of Dalal & Triggs. On the Daimler Chrysler pedestrian dataset IKSVM gives comparable accuracy to the best results (based on quadratic SVM), while being 15× faster. In these experiments our approximate IKSVM is up to 2000× faster than a standard implementation and requires 200× less memory. Finally we show that a 50× speedup is possible using approximate IKSVM based on spatial pyramid features on the Caltech 101 dataset with negligible loss of accuracy.

42. Looking around the Backyard Helps to Recognize Faces and Digits

Honghao Shan, Garrison Cottrell

Human beings have the ability to learn to recognize a new visual category based on only one or few training examples. Part of this ability might come from the use of knowledge from previous visual experiences. We show that such knowledge can be expressed as a set of "universal" visual features, which are learned from randomly collected natural scene images. Using these visual features, we have obtained state-of-the-art performance on several classification tasks using a single-layer classifier.

43. Keywords to Visual Categories: Multiple-Instance Learning for Weakly Supervised Object Categorization

Sudheendra Vijayanarasimhan, Kristen Grauman

Conventional supervised methods for image categorization rely on manually annotated (labeled) examples to learn good object models, which means their generality and scalability depends heavily on the amount of human effort available to help train them. We propose an unsupervised approach to construct discriminative models for categories specified simply by their names. We show that multiple-instance learning enables the recovery of robust category models from images returned by keyword-based search engines. By incorporating constraints that reflect the expected sparsity of true positive examples into a large-margin objective function, our approach remains accurate even when the available text annotations are imperfect and ambiguous. In addition, we show how to iteratively improve the learned classifier by automatically refining the representation of the ambiguously labeled examples. We demonstrate our method with benchmark datasets, and show that it performs well relative to both state-of-the-art unsupervised approaches and traditional fully supervised techniques.

44. Small codes and large databases of images for object recognition

Antonio Torralba, Yair Weiss, Rob Fergus

The Internet contains billions of images, freely available online. Methods for efficiently searching this incredibly rich resource are vital for a large number of applications. These include object recognition [2], computer graphics [11, 27], personal photo collections, online image search tools. In this paper, our goal is to develop efficient image search and scene matching techniques that are not only fast, but also require very little memory, enabling their use on standard hardware or even on handheld devices. Our approach uses recently developed machine learning techniques to convert the Gist descriptor (a real valued vector that describes orientation energies at different scales and orientations within an image) to a compact binary code, with a few hundred bits per image. Using our scheme, it is possible to perform real-time searches with millions from the Internet using a single large PC and obtain recognition results comparable to the full descriptor. Using our codes on high quality labeled images from the LabelMe database gives surprisingly powerful recognition results using simple nearest neighbor techniques.

45. A robust identification approach to gait recognition

Tao Ding

In this paper we address the problem of human gait recognition from a robust identification and model (in)validation prospective. The main idea is to apply dimensionality reduction technique to extract the spatio-temporal information by mapping the gait silhouette sequence to a low dimensional time sequence, which is considered as the output of a linear time invariant (LTI) system. A class of gaits is associated to a nominal discrete LTI system which has a periodic impulse response and is identified by robust identification approach. Correspondingly, gait recognition can be formulated as measuring the difference between the models representing different gait sequences. Our approach provides an efficient way to extract, to model shape-motion information of gait sequence, and to measure the difference between gait sequence models which is robust to gait cycle localization, gross appearance variation, and time scaling. These results are illustrated with practical examples on popular gait databases.

46. Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases

James Philbin, Ondrej Chum, Josef Sivic, Michael Isard, Andrew Zisserman

The state of the art in visual object retrieval from large databases is achieved by systems that are inspired by text retrieval. A key component of these approaches is that local regions of images are characterized using high-dimensional descriptors which are then mapped to “visual words” selected from a discrete vocabulary.

This paper explores techniques to map each visual region to a weighted set of words, allowing the inclusion of features which were lost in the quantization stage of previous systems. The set of visual words is obtained by selecting words based on proximity in descriptor space. We describe how this representation may be incorporated into a standard tf-idf architecture, and how spatial verification is modified in the case of this soft-assignment.

We evaluate our method on the standard Oxford Buildings dataset, and introduce a new dataset for evaluation. Our results exceed the current state of the art retrieval performance on these datasets, particularly on queries with poor initial recall where techniques like query expansion suffer. Overall we show that soft-assignment is always beneficial for retrieval with large vocabularies, at a cost of increased storage requirements for the index.

47. Fast Kernel Learning for Spatial Pyramid Matching

Junfeng He, Shih-Fu Chang, Lexing Xie

Spatial pyramid matching (SPM) is a simple yet effective approach to compute similarity between images. Similarity kernels at different regions and scales are usually fused by some heuristic weights. In this paper,we develop a novel and fast approach to improve SPM by finding the optimal kernel fusing weights from multiple scales, locations, as well as codebooks. One unique contribution of our approach is the novel formulation of kernel matrix learning problem leading to an efficient quadratic programming solution, with much lower complexity than those associated with existing solutions (e.g., semidefinite programming). We demonstrate performance gains of the proposed methods by evaluations over well-known public data sets such as natural scenes and TRECVID 2007.

48. Transfer Learning for Image Classification with Sparse Prototype Representations

Ariadna Quattoni, Michael Collins, Trevor Darrell

To learn a new visual category from few examples, prior knowledge from unlabeled data as well as previous related categories may be useful. We develop a new method for transfer learning which exploits available unlabeled data and an arbitrary kernel function; we form a representation based on kernel distances to a large set of unlabeled data points. To transfer knowledge from previous related problems we observe that a category might be learnable using only a small subset of reference prototypes. Related problems may share a significant number of relevant prototypes; we find such a concise representation by performing a joint loss minimization over the training sets of related problems with a shared regularization penalty that minimizes the total number of prototypes involved in the approximation. This optimization problem can be formulated as a linear program that can be solved efficiently. We conduct experiments on a news-topic prediction task where the goal is to predict whether an image belongs to a particular news topic. Our results show that when only few examples are available for training a target topic, leveraging knowledge learnt from other topics can significantly improve performance.

49. Optimized KD-trees for fast image descriptor matching

Chanop Silpa-Anan, Richard Hartley

In this paper, we look at improving the KD-tree for a specific usage: indexing a large number of SIFT and other types of image descriptors. We have extended priority search, to priority search among multiple trees. By creating multiple KD-trees from the same data set and simultaneously searching among these trees, we have improved the KD-tree's search performance significantly. We have also exploited the structure in SIFT descriptors (or structure in any data set) to reduce the time spent in backtracking. By using Principal Component Analysis to align the principal axes of the data with the coordinate axes, we have further increased the KD-tree's search performance.

50. An importance sampling approach to learning structural representations of shape

Andrea Torsello

This paper addresses the problem of learning archetypal structural models from examples. This is done by providing a generative model for graphs where the distribution of observed nodes and edges is governed by a set of independent Bernoulli trials with parameters to be estimated, however, the correspondences between sample node and model nodes is not known and must be estimated from local structure. The parameters are estimated maximizing the likelihood of the observed graphs, marginalizing it over all possible node correspondences. This is done adopting an importance sampling approach to limit the exponential explosion of the set of correspondences. The approach is used to summarize the variation in two different structural abstraction of shape: Delaunay graph over a set of image features and shock graphs. The experiments show that the approach can be used to recognize structures belonging to a same class.

51. Simultaneous Image Transformation and Sparse Representation Recovery

Junzhou Huang, Xiaolei Huang, Dimitris Metaxas

Sparse representation in compressive sensing is gaining increasing attention due to its success in various applications. As we demonstrate in this paper, however, image sparse representation is sensitive to image plane transformations such that existing approaches can not reconstruct the sparse representation of a geometrically transformed image. We introduce a simple technique for obtaining transformation-invariant image sparse representation. It is rooted in two observations: 1) if the aligned model images of an object span a linear subspace, their transformed versions with respect to some group of transformations can still span a linear subspace in a higher dimension; 2) if a target (or test) image, aligned with the model images, lives in the above subspace, its pre-alignment versions would get closer to the subspace after applying estimated transformations with more and more accurate parameters. These observations motivate us to project a potentially unaligned target image to random projection manifolds defined by the model images and the transformation model. Each projection is then separated into the aligned projection target and a residue due to misalignment. The desired aligned projection target is then iteratively optimized by gradually diminishing the residue. In this framework, we can simultaneously recover the sparse representation of a target image and the image plane transformation between the target and the model images. We have applied the proposed methodology to two applications: face recognition, and dynamic texture registration. The improved performance over previous methods that we obtain demonstrates the effectiveness of the proposed approach.

52. Kernel Integral Images: A Framework for Fast Non-Uniform Filtering

Mohamed Hussein, Porikli Fatih, Davis Larry

Integral images are commonly used in computer vision and computer graphics applications. Evaluation of box filters via integral images can be performed in constant time, regardless of the filter size. Although Heckbert extended the integral image approach for more complex filters, its usage has been very limited, in practice. In this paper, we present an extension to integral images that allows for application of a wide class of non-uniform filters. Our approach is superior to Heckbert's in terms of precision requirements and suitability for parallelization. We explain the theoretical basis of the approach and instantiate two concrete examples: filtering with bilinear interpolation, and filtering with approximated Gaussian weighting. Our experiments show the significant speedups we achieve, and the higher accuracy of our approach compared to Heckbert's.

53. The patch transform and its applications to image editing

Taeg Sang Cho, Moshe Butman, Shai Avidan, William T. Freeman

We introduce the patch transform, where an image is broken into non-overlapping patches, and modifications or constraints are applied in the “patch domain”. A modified image is then reconstructed from the patches, subject to those constraints. When no constraints are given, the reconstruction problem reduces to solving a jigsaw puzzle. Constraints the user may specify include the spatial locations of patches, the size of the output image, or the pool of patches from which an image is reconstructed. We define terms in a Markov network to specify a good image reconstruction from patches: neighboring patches must fit to form a plausible image, and each patch should be used only once. We find an approximate solution to the Markov network using loopy belief propagation, introducing an approximation to handle the combinatorially difficult patch exclusion constraint. The resulting image reconstructions show the original image, modified to respect the user's changes. We apply the patch transform to various image editing tasks and show that the algorithm performs well on real world images.

54. Visibility in Bad Weather from A Single Image

Robby Tan

Bad weather, such as fog and haze, can significantly degrade the visibility of a scene. Optically, this is due to the substantial presence of particles in the atmosphere that absorb and scatter light. In computer vision, the absorption and scattering processes are commonly modeled by a linear combination of the direct attenuation and the airlight. Based on this model, a few methods have been proposed, and most of them require multiple input images of a scene, which have either different degrees of polarization or different atmospheric conditions. This requirement is the main drawback of these methods, since in many situations, it is difficult to be fulfilled. To resolve the problem, we introduce an automated method that only requires a single input image. This method is based on two basic observations: first, images with enhanced visibility (or clear-day images) have more contrast than images plagued by bad weather; second, airlight whose variation mainly depends on the distance of objects to the viewer, tends to be smooth. Relying on these two observations, we develop a cost function in the framework of Markov random fields, which can be efficiently optimized by various techniques, such as graph-cuts or belief propagation. The method does not require the geometrical information of the input image, and is applicable for both color and gray images.

55. A Robust Descriptor based on Weber's Law

Jie Chen, Shiguang Shan, Guoying Zhao, Xilin Chen, Wen Gao, Matti Pietikainen

Inspired by Weber's Law, this paper proposes a simple, yet very powerful and robust local descriptor, Weber Local Descriptor (WLD). It is based on the fact that human perception of a pattern depends on not only the change of a stimulus (such as sound, lighting, et al.) but also the original intensity of the stimulus. Specifically, WLD consists of two components: its differential excitation and orientation. A differential excitation is a function of the ratio between two terms: One is the relative intensity differences of its neighbors against a current pixel; the other is the intensity of the current pixel. An orientation is the gradient orientation of the current pixel. For a given image, we use the differential excitation and the orientation components to construct a concatenated WLD histogram feature. Experimental results on Brodatz textures show that WLD impressively outperforms the other classical descriptors (e.g., Gabor). Especially, experimental results on face detection show a promising performance. Although we train only one classifier based on WLD features, the classifier obtains a comparable performance to state-of-the-art methods on MIT+CMU frontal face test set, AR face dataset and CMU profile test set.

56. Boosting Ordinal Features for Accurate and Fast Iris Recognition

Zhaofeng He, Zhenan Sun, Tieniu Tan, Xianchao Qiu, Cheng Zhong, Wenbo Dong

In this paper, we present a novel iris recognition method based on learned ordinal features. Firstly, taking full advantages of the properties of iris textures, a new iris representation method based on regional ordinal measure encoding is presented, which provides an over-complete iris feature set for learning. Secondly, a novel Similarity Oriented Boosting (SOBoost) algorithm is proposed to train an efficient and stable classifier with a small set of features. Compared with Adaboost, SOBoost is advantageous in that it operates on similarity oriented training samples, and therefore provides a better way for boosting strong classifiers. Finally, the well-known cascade architecture is adopted to reorganize the learned SOBoost classifier into a 'cascade', by which the searching ability of iris recognition towards large-scale deployments is greatly enhanced. Extensive experiments on two challenging iris image databases demonstrate that the proposed method achieves state-of-the-art iris recognition accuracy and speed. In addition, SOBoost outperforms Adaboost (Gentle-Adaboost, JS-Adaboost, etc.) in terms of both accuracy and generalization capability across different iris databases.

57. Robust 3D Face Recognition in Un-Controlled Environments

Cheng Zhong, Zhenan Sun, Tieniu Tan, Zhaofeng He

Most current 3D face recognition algorithms are designed based on the data collected in controlled situations, which leads to the un-guaranteed performance in practical systems. In this paper, we propose a Robust Local Log-Gabor Histograms (RLLGH) method to handle the uncontrolled problems encountered in 3D face recognition. In this challenging topic, large expressions and data noises are two main obstacles. To overcome the large expressions, we choose Log-Gabor features (LGF) to extract the distinctive and robust information embedded in 3D faces, which will be represented as 3D Log-Gabor faces. Data noises are summarized as distorted meshes, hair occlusions and misalignments. To overcome these problems, we introduce a Robust Local Histogram (RLH) strategy, which takes advantage of the robustness of the accurate local statistical information. The combination of LGF and RLH leads to RLLGH. The novelties of this paper come from 1) Our work aims at studying 3D face recognition performance in uncontrolled environments; 2) We find that embedding LGF into the LVC framework leads to robustness in handling large expression variations; 3) The RLH strategy gives a promising way to solve the data noises problem. Our experiments are based on the large expression subset in FRGC2.0 3D face database and the expression subset in CASIA 3D face database. Experimental results show the efficiency, robustness and generalization of our proposed method.

58. Image Super-Resolution as Sparse Representation of Raw Image Patches

Jianchao Yang, John Wright, Thomas Huang, Ma Yi

This paper addresses the problem of generating a super-resolution (SR) image from a single low-resolution input image. We approach this problem from the perspective of compressed sensing. The low-resolution image is viewed as downsampled version of a high-resolution image, whose patches are assumed to have a sparse representation with respect to an over-complete dictionary of prototype signal-atoms. The principle of compressed sensing ensures that under mild conditions, the sparse representation can be correctly recovered from the downsampled signal. We will demonstrate the effectiveness of sparsity as a prior for regularizing the otherwise ill-posed super-resolution problem. We further show that a small set of randomly chosen raw patches from training images of similar statistical nature to the input image generally serve as a good dictionary, in the sense that the computed representation is sparse and the recovered high-resolution image is competitive or even superior in quality to images produced by other SR methods.

59. Radiometric Calibration with Illumination Change for Outdoor Scene Analysis

Seon Joo Kim, Jan-Michael Frahm, Marc Pollefeys

The images of an outdoor scene collected over time are valuable in studying the scene appearance variation which can lead to novel applications and help enhance existing methods that were constrained to controlled environments. However, the images do not reflect the true appearance of the scene in many cases due to the radiometric properties of the camera : the radiometric response function and the changing exposure. We introduce a new algorithm to compute the radiometric response function and the exposure of images given a sequence of images of a static outdoor scene where the illumination is changing. We use groups of pixels with constant behaviors towards the illumination change for the response estimation and introduce a sinusoidal lighting variation model representing the daily motion of the sun to compute the exposures.

60. Image Decomposition into Structure and Texture Subcomponents with Multifrequency Modulation Constraints

Georgios Evangelopoulos, Petros Maragos

Texture information in images is coupled with geometric macrostructures and piecewise-smooth intensity variations. Decomposing an image f into a geometric structure component u and a texture component v is an inverse estimation problem, essential for understanding and analyzing images depending on their content. In this paper, we present a novel combined approach for simultaneous texture from structure separation and multiband texture modeling. First, we formulate a new, variational decomposition scheme, involving an explicit texture reconstruction constraint (prior) formed by the responses of selected frequency-tuned linear filters. This forms a '$u + Kv$' image model of $K + 1$ components. Subsequent texture modeling is applied to the estimated v component and its consistency is compared to using the complete, initial image f. The decomposition step, functioning as an advanced texture-front end, improves clustering and classification performance, for various multiband features. The proposed method can be generalized to other texture models or applications.

61. Radiometric Calibration Using Temporal Irradiance Mixtures

Bennett Wilburn, Hui Xu, Yasuyuki Matsushita

We propose a new method for sampling camera response functions: temporally mixing two uncalibrated irradiances within a single camera exposure. Calibration methods rely on some known relationship between irradiance at the camera image plane and measured pixel intensities. Prior approaches use a color checker chart with known reflectances, registered images with different exposure ratios, or even the irradiance distribution along edges in images. We show that temporally blending irradiances allows us to densely sample the camera response function with known relative irradiances. Our first method computes the camera response curve using temporal mixtures of two pixel intensities on an uncalibrated computer display. The second approach makes use of temporal irradiance mixtures caused by motion blur. Both methods require only one input image, although more images can be used for improved robustness to noise or to cover more of the response curve. We show that our methods compute accurate response functions for a variety of cameras.

62. Light-Invariant Fitting of Active Appearance Models

Daniel Pizarro, Julien Peyras, Adrien Bartoli

This paper deals with shading and AAMs. Shading is created by lighting change. It can be of two types: self-shading and external shading. The effect of self-shading can be explicitly learned and handled by AAMs. This is not however possible for external shading, which is usually dealt with by robustifying the cost function.

We take a different approach: we measure the fitting cost in a so-called Light-Invariant space. This approach naturally handles self-shading and external shading. The framework is based on mild assumptions on the scene reflectance and the cameras. Some photometric camera response parameters are required. We propose to estimate these while fitting an existing color AAM in a photometric `self-calibration' manner.

We report successful results with a face AAM with test images taken indoor under simple lighting change.

63. Discriminative Learned Dictionaries for Local Image Analysis

Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, Andrew Zisserman

Sparse signal models have been the focus of much recent research, leading to (or improving upon) state-of-the-art results in signal, image, and video restoration. This article extends this line of research into a novel framework for local image discrimination tasks, proposing an energy formulation with both sparse reconstruction and class discrimination components, jointly optimized during dictionary learning. This approach improves over the state of the art in texture segmentation experiments using the Brodatz database, and it paves the way for a novel scene analysis and recognition framework based on simultaneously learning discriminative and reconstructive dictionaries. Preliminary results in this direction using examples from the Pascal VOC06 and Graz02 datasets are presented as well.

64. Demosaicing by Smoothing along 1D Features

Boris Ajdin, Matthias B. Hullin, Christian Fuchs, Hans-Peter Seidel, Hendrik P. A. Lensch

Most digital cameras capture color pictures in the form of an image mosaic, recording only one color channel at each pixel position. Therefore, an interpolation algorithm needs to be applied to reconstruct the missing color information. In this paper we present a novel Bayer pattern demosaicing approach, employing stochastic global optimization performed on a pixel neighborhood. We are minimizing a newly developed cost function that increases smoothness along one-dimensional image features. While previous algorithms have been developed focusing on LDR images only, our optimization scheme and the underlying cost function are designed to handle both LDR and HDR images, creating less demosaicing artifacts, compared to previous approaches.

65. Histogram-based search: A comparative study

Mikhail Sizintsev, Konstantinos G. Derpanis, Andrew Hogue

Histograms represent a popular means for feature representation. This paper is concerned with the problem of exhaustive histogram-based image search. Several standard histogram construction methods are explored, including the conventional approach, Huang's method, and the state-of-the-art integral histogram. In addition, we present a novel multiscale histogram-based search algorithm, termed the distributive histogram, that can be evaluated exhaustively in a fast and memory efficient manner. An extensive systematic empirical evaluation is presented that explores the computational and storage consequences of altering the search image and histogram bin sizes. Experiments reveal up to an eight-fold decrease in computation time and hundreds- to thousands-fold decrease of memory use of the proposed distributive histogram in comparison to the integral histogram. Finally, we conclude with a discussion on the relative merits between the various approaches considered in the paper.

66. Estimating Camera Response Functions using Probabilistic Intensity Similarity

Jun Takamatsu, Yasuyuki Matsushita, Katsushi Ikeuchi

We propose a method for estimating camera response functions using a probabilistic intensity similarity measure. The similarity measure represents the likelihood of two intensity observations corresponding to the same scene radiance in the presence of noise. We show that the response function and the intensity similarity measure are strongly related. Our method requires several input images of a static scene taken from the same viewing position with fixed camera parameters. Noise causes pixel values at the same pixel coordinate to vary in these images, even though they measure the same scene radiance. We use these fluctuations to estimate the response function by maximizing the intensity similarity function for all pixels. Unlike prior noise-based estimation methods, our method requires only a small number of images, so it works with digital cameras as well as video cameras. Moreover, our method does not rely on any special image processing or statistical prior models. Real-world experiments using different cameras demonstrate the effectiveness of the technique.

67. Photometric stereo with non-parametric and spatially-varying reflectance

Neil Alldrin, Todd Zickler, David Kriegman

We present a method for simultaneously recovering shape and spatially varying reflectance of a surface from photometric stereo images. The distinguishing feature of our approach is its generality; it does not rely on a specific parametric reflectance model and is therefore purely “data-driven”. This is achieved by employing novel bi-variate approximations of isotropic reflectance functions. By combining this new approximation with recent developments in photometric stereo, we are able to simultaneously estimate an independent surface normal at each point, a global set of non-parametric “basis material” BRDFs, and per-point material weights. Our experimental results validate the approach and demonstrate the utility of bi-variate reflectance functions for general non-parametric appearance capture.

68. Demosaicking Recognition with Applications in Digital Photo Authentication based on a Quadratic Pixel Correlation Model

Yizhen Huang, Jingying Wen, Yangjing Long

Most digital still color cameras use a single electronic sensor (CCD or CMOS) overlaid with a color filter array. At each pixel location only one color sample is taken, and the other colors must be interpolated using neighboring samples. This color plane interpolation is known as demosaicking, which is one of the important tasks in a digital camera pipeline. Demosaicked images possess spatially periodic inter-pixel correlation. In this paper, such correlation is expressed in a quadratic form, and Principal Component Analysis is applied to filter out intrinsic scene correlation. A decision mechanism using BP neural networks and a majority-voting scheme is designed to recognize demosaicking correlation and authenticate digital photos. Experiments show that, the proposed method can accurately classify images by demosaicking algorithms or source cameras, and it is effective to detect rendering forgeries. The sensitivity and robustness of the method are also verified. This algorithm-independent approach is especially useful when demosaicking algorithm is only available in form of binary code or integrated circuit without technical detail.

69. Evaluation of Color Descriptors for Object and Scene Recognition

Koen E. A. van de Sande, Theo Gevers, Cees G.M. Snoek

Image category recognition is important to access visual information on the level of objects and scene types. So far, intensity-based descriptors have been widely used. To increase illumination invariance and discriminative power, color descriptors have been proposed only recently. As many descriptors exist, a structured overview of color invariant descriptors in the context of image category recognition is required. Therefore, this paper studies the invariance properties and the distinctiveness of color descriptors in a structured way. The invariance properties of color descriptors are shown analytically using a taxonomy based on invariance properties with respect to photometric transformations. The distinctiveness of color descriptors is assessed experimentally using two benchmarks from the image domain and the video domain. From the theoretical and experimental results, it can be derived that invariance to light intensity changes and light color changes affects category recognition. The results reveal further that, for light intensity changes, the usefulness of invariance is category-specific.

70. Image Super-Resolution using Gradient Profile Prior

Jian Sun, Jian Sun, Xu Zongben, Shum Heung-Yeung

In this paper, we propose an image super-resolution approach using a novel generic image prior -- gradient profile prior, which is a parametric prior describing the shape and the sharpness of the image gradients. Using the gradient profile prior learned from a large number of natural images, we can provide a constraint on image gradients when we estimate a hi-resolution image from a low-resolution image. With this simple but very effective prior, we are able to produce state-of-the-art results. The reconstructed high-resolution image is sharp while has rare ringing or jaggy artifacts.

71. Intrinsic image decomposition with non-local texture cues

Li Shen, Ping Tan, Stephen Lin

We present a method for decomposing an image into its intrinsic reflectance and shading components. Different from previous work, our method examines texture information to obtain constraints on reflectance among pixels that may be distant from one another in the image. We observe that distinct points with the same intensity-normalized texture configuration generally have the same reflectance value. The separation of shading and reflectance components should thus be performed in a manner that guarantees these non-local constraints. We formulate intrinsic image decomposition by adding these non-local texture constraints to the local derivative analysis employed in conventional techniques. Our results show a significant improvement in performance, with better recovery of global reflectance and shading structure than by previous methods.

72. Automatic registration of aerial imagery with untextured 3D LiDAR model

Min Ding, Kristian Lyngbaek, Avideh Zakhor

A fast 3D model reconstruction methodology is desirable in many applications such as urban planning, training, and simulations. In this paper, we develop an automated algorithm for texture mapping oblique aerial images onto a 3D model generated from airborne Light Detection and Ranging (LiDAR) data. Our proposed system consists of two steps. In the first step, we combine vanishing points and global positioning system aided inertial system readings to roughly estimate the extrinsic parameters of a calibrated camera. In the second step, we refine the coarse estimate of the first step by applying a series of processing steps. Specifically, We extract 2D corners corresponding to orthogonal 3D structural corners as features from both images and the untextured 3D LiDAR model. The correspondence between an image and the 3D model is then performed using Hough transform and generalized M-estimator sample consensus. The resulting 2D corner matches are used in Lowe's algorithm to refine camera parameters obtained earlier. Our system achieves 91% correct pose recovery rate for 90 images over the downtown Berkeley area, and overall 61% accuracy rate for 358 images over the residential, downtown and campus portions of the city of Berkeley.

73. Approximate earth mover's distance in linear time

Sameer Shirdhonkar, David W. Jacobs

The earth mover's distance (EMD) is an important perceptually meaningful metric for comparing histograms, but it suffers from high ($O(N^3 \log N)$) computational complexity. We present a novel linear time algorithm for approximating the EMD for low dimensional histograms using the sum of absolute values of the weighted wavelet coefficients of the difference histogram. EMD computation is a special case of the Kantorovich-Rubinstein transshipment problem, and we exploit the Hölder continuity constraint in its dual form to convert it into a simple optimization problem with an explicit solution in the wavelet domain. We prove that the resulting wavelet EMD metric is equivalent to EMD, i.e., the ratio of the two is bounded. We also provide estimates for the bounds.

The weighted wavelet transform can be computed in time linear in the number of histogram bins, while the comparison is about as fast as for normal Euclidean distance or χ^2 statistic. We experimentally show that wavelet EMD is a good approximation to EMD, has similar performance, but requires much less computation.

74. Texture Classification with a Dictionary of Basic Image Features

Michael Crosier, Lewis D. Griffin

Many successful recent approaches to texture classification model texture images as distributions over a set of discrete features, or textons, which correspond to a partitioning of the space of responses to local descriptors such as filter banks or image patches. This partitioning is learned by unsupervised clustering of descriptor responses taken from the dataset to be analysed. Here, we explore a quantization of filter responses into a dictionary of discrete features which is based on geometrical, rather than statistical, considerations, resulting in a simple texture description based on a dictionary of 'visual words' which is independent of the images to be described. A multi-scale classification scheme built on this dictionary is evaluated. The results presented are, to the best of our knowledge, state-of-the-art for the UIUCTex and KTH-TIPS datasets, and close to the state-of-the-art for CUReT, despite using a less sophisticated classifier.

75. Color constancy beyond bags of pixels

Ayan Chakrabarti, Keigo Hirakawa, Todd Zickler

Estimating the color of a scene illuminant often plays a central role in computational color constancy. While this problem has received significant attention, the methods that exist do not maximally leverage spatial dependencies between pixels. Indeed, most methods treat the observed color (or its spatial derivative) at each pixel independently of its neighbors. We propose an alternative approach to illuminant estimation--one that employs an explicit statistical model to capture the spatial dependencies between pixels induced by the surfaces they observe. The parameters of this model are estimated from a training set of natural images captured under canonical illumination, and for a new image, an appropriate transform is found such that the corrected image best fits our model.

3:45pm – 5:51pm Oral Session O2P-1: Statistical Methods and Visual Learning (Cook)

1. Non-Negative Graph Embedding

Jianchao Yang, Shuicheng Yan, Yun Fu, Xuelong Li, Thomas Huang

We introduce a general formulation, called non-negative graph embedding, for non-negative data decomposition by integrating the characteristics of both intrinsic and penalty graphs. In the past, such a decomposition was obtained mostly in an unsupervised manner, such as Non-negative Matrix Factorization (NMF) and its variants, and hence unnecessary to be powerful at classification. In this work, the non-negative data decomposition is studied in a unified way applicable for both unsupervised and supervised/semi-supervised configurations. The ultimate data decomposition is separated into two parts, which separatively preserve the similarities measured by the intrinsic and penalty graphs, and together minimize the data reconstruction error. An iterative procedure is derived for such a purpose, and the algorithmic non-negativity is guaranteed by the non-negative property of the inverse of any *M*-matrix. Extensive experiments compared with NMF and conventional solutions for graph embedding demonstrate the algorithmic properties in sparsity, classification power, and robustness to image occlusions.

2. Dimensionality Reduction by Unsupervised Regression

Miguel Á. Carreira-Perpiñán, Zhengdong Lu

We consider the problem of dimensionality reduction, where given high-dimensional data we want to estimate two mappings: from high to low dimension (dimensionality reduction) and from low to high dimension (reconstruction). We adopt an unsupervised regression point of view by introducing the unknown low-dimensional coordinates of the data as parameters, and formulate a regularised objective functional of the mappings and low-dimensional coordinates. Alternating minimisation of this functional is straightforward: for fixed low-dimensional coordinates, the mappings have a unique solution; and for fixed mappings, the coordinates can be obtained by finite-dimensional nonlinear minimisation. Besides, the coordinates can be initialised to the output of a spectral method such as Laplacian eigenmaps. The model generalises PCA and several recent methods that learn one of the two mappings but not both; and, unlike spectral methods, our model provides out-of-sample mappings by construction. Experiments with toy and real-world problems show that the model is able to learn mappings for convoluted manifolds, avoiding bad local optima that plague other methods.

3. Directional Independent Component Analysis with Tensor Representation

Lei Zhang, Quanxue Gao, David Zhang

Conventional independent component analysis (ICA) learns the statistical independencies of 2D variables from the training images that are unfolded to vectors. The unfolded vectors, however, make the ICA suffer from the small sample size (SSS) problem that leads to the dimensionality dilemma. This paper presents a novel directional multilinear ICA method to solve those problems by encoding the input image or high dimensional data array as a general tensor. In addition, the mode-k matrix of the tensor is re-sampled and re-arranged to form a mode-k directional image to better exploit the directional information in training. An algorithm called mode-k directional ICA is then presented for feature extraction. Compared with the conventional ICA and other subspace analysis algorithms, the proposed method can greatly alleviate the SSS problem, reduce the computational cost in the learning stage by representing the data in lower dimension, and simultaneously exploit the directional information in the high dimensional dataset. Experimental results on well-known face and palmprint databases show that the proposed method has higher recognition accuracy than many existing ICA, PCA and even supervised FLD schemes while using a low dimension of features.

4. Order Consistent Change Detection via Fast Statistical Significance Testing

Maneesh Singh, Vasu Parameswaran, Visvanathan Ramesh

Robustness to illumination variations is a key requirement for the problem of change detection which in turn is a fundamental building block for many visual surveillance applications. The use of ordinal measures is a powerful way of filtering out illumination dependency in representing appearance, and several such measures have been proposed in the past for change detection. By design, these measures are invariant to unknown monotonic transformations that may be caused due to global illumination changes or automatic camera gain. However, previous work has left theoretical and practical gaps that limit their full potential from being realized. For instance, random noise has not been given a principled treatment. In this paper, we formulate the change detection problem in terms of order consistency and show that in the presence of noise with known statistical properties, significance tests for order consistency yield much better results than the state of the art. Since ordinal measures require a reordering of patches, they are usually expensive in practice ($O(n \log n)$ at best). We improve upon this by connecting the problem to monotonic regression, and applying a fast algorithm from the corresponding literature. We also show that good trade offs between speed and accuracy can be made by quantization to achieve accurate and very fast matching algorithms in practice. We demonstrate superior performance on statistical simulations as well as real image sequences.

5. The Logistic Random Field -- A Convenient Graphical Model for Learning Parameters for MRF-based Labeling

Marshall Tappen, Kegan Samuel, Craig Dean, David Lyle

Graphical models are fundamental tools for modeling images and other applications. In this paper, we propose the Logistic Random Field (LRF) model for representing a discrete-valued graphical model. The LRF model is based on an underlying quadratic model and a logistic function. The chief advantages of the LRF are its convenience and flexibility. The quadratic model makes inference easy to implement using standard numerical linear algebra routines. This quadratic model also allows the log-likelihood of the training data to be differentiated with respect to any parameter in the model, enhancing the flexibility of the LRF model. To demonstrate the usefulness of this model we use it to learn how to segment objects, specifically roads, horses, and cows. In addition, we demonstrate the flexibility of the LRF model by incorporating super-pixels. We then show that the LRF segmentation model produces segmentations that are competitive with recently published results.

6. Large-Scale Manifold Learning

Ameet Talwalkar, Sanjiv Kumar, Henry Rowley

This paper examines the problem of extracting low-dimensional manifold structure given millions of high-dimensional face images. Specifically, we address the computational challenges of nonlinear dimensionality reduction via Isomap and Laplacian Eigenmaps, using a graph containing about 18 million nodes and 65 million edges. Since most manifold learning techniques rely on spectral decomposition, we first analyze two approximate spectral decomposition techniques for large dense matrices (Nyström and Column-sampling), providing the first direct theoretical and empirical comparison between these techniques. We next show extensive experiments on learning low-dimensional embeddings for two large face datasets: CMU-PIE (35 thousand faces) and a web dataset (18 million faces). Our comparisons show that the Nyström approximation is superior to the Column-sampling method. Furthermore, approximate Isomap tends to perform better than Laplacian Eigenmaps on both clustering and classification with the labeled CMU-PIE dataset.

3:45pm – 5:51pm Oral Session O2P-2: Stereo (Arteaga)

1. Variable Baseline/Resolution Stereo

David Gallup, Jan-Michael Frahm, Philippos Mordohai, Marc Pollefeys

We present a novel multi-baseline, multi-resolution stereo method, which varies the baseline and resolution proportionally to depth to obtain a reconstruction in which the depth error is constant. This is in contrast to traditional stereo, in which the error grows quadratically with depth, which means that the accuracy in the near range far exceeds that of the far range. This accuracy in the near range is unnecessarily high and comes at significant computational cost. It is, however, non-trivial to reduce this without also reducing the accuracy in the far range. Many datasets, such as video captured from a moving camera, allow the baseline to be selected with significant flexibility. By selecting an appropriate baseline and resolution (realized using an image pyramid), our algorithm computes a depthmap which has these properties: 1) the depth accuracy is constant over the reconstructed volume, 2) the computational effort is spread evenly over the volume, 3) the angle of triangulation is held constant w.r.t. depth. Our approach achieves a given target accuracy with minimal computational effort, and is orders of magnitude faster than traditional stereo.

2. Global Stereo Reconstruction under Second Order Smoothness Priors

Oliver Woodford, Ian Reid, Philip H. S. Torr, Andrew Fitzgibbon

Second-order priors on the smoothness of 3D surfaces are a better model of typical scenes than first-order priors. However, stereo reconstruction using global inference algorithms, such as graph-cuts, has not been able to incorporate second-order priors because the triple cliques needed to express them yield intractable (non-submodular) optimization problems.

This paper shows that inference with triple cliques can be effectively optimized. Our optimization strategy is a development of recent extensions to α-expansion, based on the "QPBO" algorithm. The strategy is to repeatedly merge proposal depth maps using a novel extension of QPBO. Proposal depth maps can come from any source, for example fronto-parallel planes as in α-expansion, or indeed any existing stereo algorithm, with arbitrary parameter settings.

Experimental results demonstrate the usefulness of the second-order prior and the efficacy of our optimization framework. An implementation of our stereo framework is available online.

3. A Fast Local Descriptor for Dense Matching

Engin Tola, Vincent Lepetit, Pascal Fua

We introduce a novel local image descriptor designed for dense wide-baseline matching purposes. We feed our descriptors to a graph-cuts based dense depth map estimation algorithm and this yields better wide-baseline performance than the commonly used correlation windows for which the size is hard to tune. As a result, unlike competing techniques that require many high-resolution images to produce good reconstructions, our descriptor can compute them from pairs of low-quality images such as the ones captured by video streams.

Our descriptor is inspired from earlier ones such as SIFT and GLOH but can be computed much faster for our purposes. Unlike SURF which can also be computed efficiently at every pixel, it does not introduce artifacts that degrade the matching performance.

Our approach was tested with ground truth laser scanned depth maps as well as on a wide variety of image pairs of different resolutions and we show that good reconstructions are achieved even with only two low quality images.

4. Fast and Robust Numerical Solutions to Minimal Problems for Cameras with Radial Distortion

Martin Byröd, Zuzana Kukelova, Klas Josephson, Tomas Pajdla, Karl Åström

A number of minimal problems of structure from motion for cameras with radial distortion have recently been studied and solved in some cases. These problems are known to be numerically very challenging and in several cases there exist no known practical algorithm yielding solutions in floating point arithmetic. We make some crucial observations concerning the floating point implementation of Gröbner basis computations and use these new insights to formulate fast and stable algorithms for two minimal problems with radial distortion previously solved in exact rational arithmetic only: (i) simultaneous estimation of essential matrix and a common radial distortion parameter for two partially calibrated views and six image point correspondences and (ii) estimation of fundamental matrix and two different radial distortion parameters for two uncalibrated views and nine image point correspondences. We demonstrate on simulated and real experiments that these two problems can be efficiently solved in floating point arithmetic.

5. Spectrally Optimal Factorization of Incomplete Matrices

Pedro M. Q. Aguiar, Joao M. F. Xavier, Marko Stosic

From the recovery of structure from motion to the separation of style and content, many problems in computer vision have been successfully approached by using bilinear models. The reason for the success of these models is that a globally optimal decomposition is easily obtained from the Singular Value Decomposition (SVD) of the observation matrix. However, in practice, the observation matrix is often incomplete, the SVD can not be used, and only suboptimal solutions are available. The majority of these solutions are based on iterative local refinements of a given cost function, and lack any guarantee of convergence to the global optimum. In this paper, we propose a globally optimal solution, for particular patterns of missing entries. To achieve this goal, we re-formulate the problem as the minimization of the spectral norm of the matrix of residuals, i.e., we seek the completion of the observation matrix such that the largest singular value of its difference to a low rank matrix is the smallest possible. The class of patterns of missing entries we deal with is known as the Young diagram, which includes, as particular cases, many relevant situations, such as the missing of an entire submatrix. We describe experiments that illustrate how our globally optimal solution has impact in practice.

6. Simple calibration of non-overlapping cameras with a mirror

Ram Krishan Kumar, Adrian Ilie, Jan Frahm, Marc Pollefeys

Calibrating a network of cameras with non-overlapping views is an important and challenging problem in computer vision. In this paper, we present a novel technique for camera calibration using a planar mirror. We overcome the need for all cameras to see a common calibration object directly by allowing them to see it through a mirror. We use the fact that the mirrored views generate a family of mirrored camera poses that uniquely describe the real camera pose. Our method consists of the following two steps: (1) using standard calibration methods to find the internal and external parameters of a set of mirrored camera poses, (2) estimating the external parameters of the real cameras from their mirrored poses by formulating constraints between them. We demonstrate our method on real and synthetic data for camera clusters with small overlap between the views and non-overlapping views.

8:30am – 10:30am Poster Session P3A-1: Stereo, Structure from Motion, Image and Video Retrieval, Object Detection and Categorization (Summit Hall)

1. Classification and evaluation of cost aggregation methods for stereo correspondence

Federico Tombari, Stefano Mattoccia, Luigi Di Stefano, Elisa Addimanda

In the last decades several cost aggregation methods aimed at improving the robustness of stereo correspondence within local and global algorithms have been proposed. Given the recent developments and the lack of an appropriate comparison, in this paper we survey, classify and compare experimentally on a standard data set the main cost aggregation approaches proposed in literature. The experimental evaluation addresses both accuracy and computational requirements, so as to outline the best performing methods under these two criteria.

2. Skeletal graphs for efficient structure from motion

Noah Snavely, Steven M. Seitz, Richard Szeliski

We address the problem of efficient structure from motion for large, unordered, highly redundant, and irregularly sampled photo collections, such as those found on Internet photo-sharing sites. Our approach computes a small *skeletal* subset of images, reconstructs the skeletal set, and adds the remaining images using pose estimation. Our technique drastically reduces the number of parameters that are considered, resulting in dramatic speedups, while provably approximating the covariance of the full set of parameters. To compute a skeletal image set, we first estimate the accuracy of two-frame reconstructions between pairs of overlapping images, then use a graph algorithm to select a subset of images that, when reconstructed, approximates the accuracy of the full set. A final bundle adjustment can then optionally be used to restore any loss of accuracy.

3. Re-Thinking Non-Rigid Structure From Motion

Vincent Rabaud, Serge Belongie

We present a novel approach to non-rigid structure from motion (NRSFM) from an orthographic video sequence, based on a new interpretation of the problem. Existing approaches assume the object shape space is well-modeled by a linear subspace. Our approach only assumes that small neighborhoods of shapes are well-modeled with a linear subspace. This constrains the shapes to belong to a manifold of dimensionality equal to the number of degrees of freedom of the object. After showing that the problem is still overconstrained, we present a solution composed of a novel initialization algorithm, followed by a robust extension of the Locally Smooth Manifold Learning algorithm tailored to the NRSFM problem. We finally present some test cases where the linear basis method fails (and is actually not meant to work) while the proposed approach is successful.

4. Motion Estimation for Multi-Camera Systems using Global Optimization

Jae-Hak Kim, Hongdong Li, Richard Hartley

We present a motion estimation algorithm for multi-camera systems consisting of more than one calibrated camera securely attached on a moving object. So, they move all together, but do not require to have overlapping views across the cameras. The geometrically optimal solution of the motion for the multi-camera systems under L_∞ norm is provided in this paper using a global optimization technique which has been introduced recently in the computer vision research field. Taking advantage of an optimal estimate of the essential matrix through searching rotation space, we provide the optimal solution for translation by using linear programming and Branch & Bound algorithm. Synthetic and real data experiments are conducted, and they show more robust and improved performance than the previous methods.

5. Accurate Calibration from Multi-View Stereo and Bundle Adjustment

Yasutaka Furukawa, Jean Ponce

The advent of high-resolution digital cameras and sophisticated multi-view stereo algorithms offers the promises of unprecedented geometric fidelity in image-based modeling tasks, but it also puts unprecedented demands on camera calibration to fulfill these promises. This paper presents a novel approach to camera calibration where top-down information from rough camera parameter estimates and the output of a publicly available multiview-stereo system on scaled-down input images are used to effectively guide the search for additional image correspondences and significantly improve camera calibration parameters using a standard bundle adjustment algorithm. The proposed method has been tested on several real datasets--including objects without salient features for which image correspondences cannot be found in a purely bottom-up fashion, and image-based modeling tasks--including the construction of visual hulls where thin structures are lost without our calibration procedure.

6. Silhouette-Based Camera Calibration from Sparse Views under Circular Motion

Po-Hao Huang, Shang-Hong Lai

In this paper, we propose a new approach to camera calibration from silhouettes under circular motion with minimal data. We exploit the mirror symmetry property and derive a common homography that relates silhouettes with epipoles under circular motion. With the epipoles determined, the homography can be computed from the frontier points induced by epipolar tangencies. On the other hand, given the homography, the epipoles can be located directly from the bi-tangent lines of silhouettes. With the homography recovered, the image invariants under circular motion and camera parameters can be determined. If the epipoles are not available, camera parameters can be determined by a low-dimensional search of the optimal homography in a bounded region. In the degenerate case, when the camera optical axes intersect at one point, we derive a closed-form solution for the focal length to solve the problem. By using the proposed algorithm, we can achieve camera calibration simply from silhouettes of three images captured under circular motion. Experimental results on synthetic and real images are presented to show its performance.

7. Dense 3D Reconstruction from Specularity Consistency

Diego Nehab, Tim Weyrich, Szymon Rusinkiewicz

In this work, we consider the dense reconstruction of specular objects. We propose the use of a specularity constraint, based on surface normal/depth consistency, to define a matching cost function that can drive standard stereo reconstruction methods. We discuss the types of ambiguity that can arise, and suggest an aggregation method based on anisotropic diffusion that is particularly suitable for this matching cost function.

We also present a controlled illumination setup that includes a pair of cameras and one LCD monitor, which is used as a calibrated, variable-position light source. We use this setup to evaluate the proposed method on real data, and demonstrate its capacity to recover high-quality depth and orientation from specular objects.

8. Edge Descriptors for Robust Wide-Baseline Correspondence

Jason Meltzer, Stefano Soatto

This paper describes a method for finding wide-baseline correspondences between images at locations along gradient edges. We find edges in scale space using established methods and develop invariant descriptors for these edges based on orientation and scale histograms. Because edges are often found on occluding boundaries, we calculate and store two descriptors per edge, one on each side, for robustness to occlusions. We demonstrate the effectiveness of edge matching in the applications of wide-baseline correspondence, structure from motion from line segments, and object category recognition on the Caltech 101 dataset.

9. Dense Specular Shape from Multiple Specular Flows

Yuriy Vasilyev, Yair Adato, Todd Zickler, Ohad Ben-Shahar

The inference of specular (mirror-like) shape is a particularly difficult problem because an image of a specular object is nothing but a distortion of the surrounding environment. Consequently, when the environment is unknown, such an image would seem to convey little information about the shape itself. It has recently been suggested (Adato et al., ICCV 2007) that observations of relative motion between a specular object and its environment can dramatically simplify the inference problem and allow one to recover shape without explicit knowledge of the environment content. However, this approach requires solving a non-linear PDE (the `shape from specular flow equation') and analytic solutions are only known to exist for very constrained motions.

In this paper, we consider the recovery of shape from specular flow under *general* motions. We show that while the `shape from specular flow' PDE for a single motion is non-linear, we can combine observations of multiple specular flows from distinct relative motions to yield a linear set of equations. We derive necessary conditions for this procedure, discuss several numerical issues with their solution, and validate our results quantitatively using image data.

10. A Three-Point Minimal Solution for Panoramic Stitching with Lens Distortion

Hailin Jin

We present a minimal solution for aligning two images taken by a rotating camera from point correspondences. The solution particularly addresses the case where there is lens distortion in the images. We assume to know the two camera centers but not the focal lengths and allow the latter to vary. Our solution uses a minimal number (three) of point correspondences and is well suited to be used in a hypothesis testing framework. It does not suffer from numerical instabilities observed in other algebraic minimal solvers and is also efficient. We validate our solution in multi-image panoramic stitching on real images with lens distortion.

11. Robust Motion Estimation and Structure Recovery from Endoscopic Image Sequences With an Adaptive Scale Kernel Consensus Estimator

Hanzi Wang, Daniel Mirota, Masaru Ishii, Gregory D. Hager

To correctly estimate the camera motion parameters and reconstruct the structure of the surrounding tissues from endoscopic image sequences, we need not only to deal with outliers (e.g., mismatches), which may involve more than 50% of the data, but also to accurately distinguish inliers (correct matches) from outliers. In this paper, we propose a new robust estimator, Adaptive Scale Kernel Consensus (ASKC), which can tolerate more than 50 percent outliers while automatically estimating the scale of inliers. With ASKC, we develop a reliable feature tracking algorithm. This, in turn, allows us to develop a complete system for estimating endoscopic camera motion and reconstructing anatomical structures from endoscopic image sequences. Preliminary experiments on endoscopic sinus imagery have achieved promising results.

12. Image Selection For Improved Multi-View Stereo

Alexander Hornung, Boyi Zeng, Leif Kobbelt

The Middlebury Multi-View Stereo evaluation clearly shows that the quality and speed of most multi-view stereo algorithms depends significantly on the number and selection of input images. In general, not all input images contribute equally to the quality of the output model, since several images may often contain similar and hence overly redundant visual information. This leads to unnecessarily increased processing times. On the other hand, a certain degree of redundancy can help to improve the reconstruction in more “difficult” regions of a model.

In this paper we propose an image selection scheme for multi-view stereo which results in improved reconstruction quality compared to uniformly distributed views. Our method is tuned towards the typical requirements of current multi-view stereo algorithms, and is based on the idea of incrementally selecting images so that the overall coverage of a simultaneously generated proxy is guaranteed without adding too much redundant information. Critical regions such as cavities are detected by an estimate of the local photo-consistency and are improved by adding additional views. Our method is highly efficient, since most computations can be out-sourced to the GPU. We evaluate our method with four different methods participating in the Middlebury benchmark and show that in each case reconstructions based on our selected images yield an improved output quality while at the same time reducing the processing time considerably.

13. Robust Unambiguous Parametrization of the Essential Manifold

Raghav Subbarao, Yakup Genc, Peter Meer

Analytic manifolds were recently used for motion averaging, segmentation and robust estimation. Here we consider the epipolar constraint for calibrated cameras, which is the most general motion model for calibrated cameras and is encoded by the essential matrix. The set of all essential matrices forms the essential manifold. We provide a theoretical characterization of the geometry of the essential manifold and develop a parametrization which associates each essential matrix with a unique point on the manifold. Our work provides a more complete theoretical analysis of the essential manifold than previous work in this direction. We show the results of using this parametrization with real data sets, while previous work concentrated on theoretical analysis with synthetic data.

14. Calibration and Rectification for Reflection Stereo

Masao Shimizu, Masatoshi Okutomi

This paper presents a calibration and rectification method for single-camera range estimation using a single complex image with a transparent plate or a double-sided half-mirror plate, which are collectively called a reflection stereo. The range to an object is obtainable by finding the correspondence on a constraint line in the complex image, which consists of a surface and a rear-surface reflected image in the transparent plate, or also includes a transmitted and internal reflected image through a double-sided half-mirror plate. The range estimation requires extrinsic parameters of the reflection stereo that include the shape and position of the plate and its refraction index. The proposed method assumes that the plate is non-parallel but planar for a local region around the point of interest in the complex image. The method estimates the extrinsic parameters from a set of displacements in the complex images. Experiments using real images demonstrate the effectiveness of the proposed calibration and rectification method along with fine range estimation results.

15. Stereo Reconstruction with Mixed Pixels Using Adaptive Over-Segmentation

Yuichi Taguchi, Bennett Wilburn, C. Lawrence Zitnick

We present an over-segmentation based, dense stereo algorithm that jointly estimates segmentation and depth. For mixed pixels on segment boundaries, the algorithm computes foreground opacity (alpha), as well as color and depth for the foreground and background. We model the scene as a collection of fronto-parallel planar segments in a reference view, and use a generative model for image formation that handles mixed pixels at segment boundaries. Our method iteratively updates the segmentation based on color, depth and shape constraints using MAP estimation. Given a segmentation, the depth estimates are updated using belief propagation. We show that our method is competitive with the state-of-the-art based on the new Middlebury stereo evaluation, and that it overcomes limitations of traditional segmentation based methods while properly handling mixed pixels. Z-keying results show the advantages of combining opacity and depth estimation.

16. Evaluation of Constructable Match Cost Measures for Stereo Correspondence Using Cluster Ranking

Daniel Neilson, Yee-Hong Yang

Stereo correspondence research often involves the comparison of techniques to determine which are better under different circumstances. The methods of comparison employed often take the form of applying the techniques to a few stereo image pairs with the technique with the lowest error rate declared superior. However, the majority of these comparisons do not contain any discussion of statistical significance; making the declared superiority of a technique statistically unreliable. In this paper we present a new evaluation method called cluster ranking that yields a statistically significant comparison of the stereo techniques being compared. Cluster ranking leverages statistical inference techniques to first rank the performance of stereo techniques on a single stereo image pair and then combine the rankings from multiple stereo pairs into an over-all ranking; in both of these rankings, only stereo techniques that are statistically different are given different ranks. We demonstrate our framework with a comparison of constructable match cost measures (those that can be assembled from a base set of components) on a data set consisting of 30 synthetic stereo pairs, with varying amounts of noise, and 18 scenes from the 2005 and 2006 Middlebury data sets. Our analysis reveals match cost measures, and measure components, that are statistically superior to all other measures depending on amount of noise, illumination, or exposure time.

17. Off-Axis Aperture Camera: 3D Shape Reconstruction and Image Restoration

Qingxu Dou, Paolo Favaro

In this paper we present a novel 3D surface and image reconstruction method based on the off-axis aperture camera. The key idea is to change the size or the 3-D location of the aperture of the camera lens so as to extract selected portions of the light field of the scene. We show that this results in an imaging device that blends defocus and stereo information, and present an image formation model that simultaneously captures both phenomena. As this model involves a non trivial deformation of the scene space, we also introduce the concept of scene space rectification and how this helps the reconstruction problem. Finally, we formulate our shape and image reconstruction problem as an energy minimization, and use a gradient flow algorithm to find the solution. Results on both real and synthetic data are shown.

18. Coarse-to-Fine Low-Rank Structure-from-Motion

Adrien Bartoli, Vincent Gay-Bellile, Umberto Castellani, Julien Peyras, Soren Olsen, Patrick Sayd

We address the problem of deformable shape and motion recovery from point correspondences in multiple perspective images. We use the low-rank shape model, i.e., the 3D shape is represented as a linear combination of unknown shape bases.

We propose a new way of looking at the low-rank shape model. Instead of considering it as a whole, we assume a coarse-to-fine ordering of the deformation modes, which can be seen as a model prior. This has several advantages. First, the high level of ambiguity of the original low-rank shape model is drastically reduced since the shape bases can not anymore be arbitrarily re-combined. Second, this allows us to propose a coarse-to-fine reconstruction algorithm which starts by computing the mean shape and iteratively adds deformation modes. It directly gives the sought after metric model, thereby avoiding the difficult upgrading step required by most of the other methods. Third, this makes it possible to automatically select the number of deformation modes as the reconstruction algorithm proceeds. We propose to incorporate two other priors, accounting for temporal and spatial smoothness, which are shown to improve the quality of the recovered model parameters.

The proposed model and reconstruction algorithm are successfully demonstrated on several videos and are shown to outperform the previously proposed algorithms.

19. Quasi-Perspective Projection with Applications to 3D Factorization from Uncalibrated Image Sequences

Guanghui Wang, Q. M. Jonathan Wu, Guoqiang Sun

The paper addresses the problem of factorization-based 3D reconstruction from uncalibrated image sequences. We propose a quasi-perspective projection model and apply the model to structure and motion recovery of rigid and nonrigid objects based on factorization of tracking matrix. The novelty and contribution of the paper lies in three aspects. First, under the assumption that the camera is far away from the object with small rotations, we propose and prove that the imaging process can be modeled by quasi-perspective projection. The model is more accurate than affine since the projective depths are implicitly embedded. Second, we apply the model to the factorization algorithm and establish the framework of rigid and nonrigid factorization under quasi-perspective assumption. Third, we propose a new and robust method to recover the transformation matrix that upgrades the factorization to the Euclidean space. The proposed method is validated and evaluated on synthetic and real image sequences and good improvements over existing solutions are observed.

20. Robust Fusion of Dynamic Shape and Normal Capture for High-quality Reconstruction of Time-varying Geometry

Naveed Ahmed, Christian Theobalt, Petar Dobrev, Hans-Peter Seidel, Sebastian Thrun

This paper describes a new passive approach to capture time-varying scene geometry in large acquisition volumes from multi-view video. It can be applied to reconstruct complete moving models of human actors that feature even slightest dynamic geometry detail, such as wrinkles and folds in clothing, and that can be viewed from 360°. Starting from multi-view video streams recorded under calibrated lighting, we first perform marker-less human motion capture based on a smooth template with no high-frequency surface detail. Subsequently, surface reflectance and time-varying normal fields are estimated based on the coarse template shape. The main contribution of this paper is a new statistical approach to solve the non-trivial problem of transforming the captured normal field that is defined over the smooth non-planar 3D template into true 3D displacements. Our spatio-temporal reconstruction method outputs displaced geometry that is accurate at each time step of video and temporally smooth, even if the input data are affected by noise.

21. Illumination and Camera Invariant Stereo Matching

Yong Seok Heo, Kyoung Mu Lee, Sang Uk Lee

Color information can be used as a basic and crucial cue for finding correspondence in a stereo matching algorithm. In a real scene, however, image colors are affected by various geometric and radiometric factors. For this reason, the raw color recorded by a camera is not a reliable cue, and the color consistency assumption is no longer valid between stereo images in real scenes. Hence the performance of most conventional stereo matching algorithms can be severely degraded under the radiometric variations. In this paper, we present a new stereo matching algorithm that is invariant to various radiometric variations between left and right images. Unlike most stereo algorithms, we explicitly employ the color formation model in our framework and propose a new measure called Adaptive Normalized Cross Correlation (ANCC) for a robust and accurate correspondence measure. ANCC is invariant to lighting geometry, illuminant color and camera parameter changes between left and right images, and does not suffer from fattening effects unlike conventional Normalized Cross Correlation (NCC). Experimental results show that our algorithm outperforms other stereo algorithms under severely different radiometric conditions between stereo images.

22. Inverse polar-ray projections for recovering projective transformations

Henry Chu, Yun Zhang

A ray projection in the inverse-polar space is proposed for recovering a projective transformation between two segmented images. The images are converted from their original Cartesian space to the inverse-polar space. Then, the two ray projections--one shift-invariant and the other shift-sensitive--of the inverse-polar images are computed to create two sets of data. Based on the obtained projection data, a two-step strategy is employed to recover the projective transformation. In the first step, the shift-invariant data are used to recover the four affine parameters. In the second step, the shift-sensitive data are used to recover the two projective parameters. The remaining two translation-related parameters are recovered in, e.g., an exhaustive search combined with the two-step recovery strategy. The proposed approach has been tested successfully to recover a variety of projective transformations between real images.

23. Learning for Stereo Vision Using the Structured Support Vector Machine

Yunpeng Li, Daniel Huttenlocher

We present a random field based model for stereo vision with explicit occlusion labeling in a probabilistic framework. The model employs non-parametric cost functions that can be learnt automatically using the structured support vector machine. The learning algorithm enables the training of models that are steered towards optimizing for a particular desired loss function, such as the metric used to evaluate the quality of the stereo labeling. Experimental results demonstrate that the performance of our method surpasses that of previous learning approaches and is comparable to the state-of-the-art for pixel-based stereo. Moreover, our method achieves good results even when trained on different image sets, in contrast with the common practice of hand tuning to specific benchmark images. In addition, we investigate the impact of graph structure on model performance. Our study shows that random field models with longer-range edges generally outperform the 4-connected grid and that this advantage is especially pronounced for noisy images.

24. Reconstructing Non-stationary Articulated Objects in Monocular Video using Silhouette Information

Saad Khan, Mubarak Shah

This paper presents an approach to reconstruct non-stationary, articulated objects from silhouettes obtained with a monocular video sequence. We introduce the concept of motion blurred scene occupancies, a direct analogy of motion blurred images but in a 3D object scene occupancy space resulting from the motion/deformation of the object. Our approach starts with an image based fusion step that combines color and silhouette information from multiple views. To this end we propose to use a novel construct: the temporal occupancy point (TOP), which is the estimated 3D scene location of a silhouette pixel and contains information about duration of time it is occupied. Instead of explicitly computing the TOP in 3D space we directly obtain it's imaged(projected) locations in each view. This enables us to handle monocular video and arbitrary camera motion in scenarios where complete camera calibration information may not be available. The result is a set of blurred scene occupancy images in the corresponding views, where the values at each pixel correspond to the fraction of total time duration that the pixel observed an occupied scene location. We then use a motion de-blurring approach to de-blur the occupancy images. The de-blurred occupancy images correspond to a silhouettes of the mean/motion compensated object shape and are used to obtain a visual hull reconstruction of the object. We show promising results on challenging monocular datasets of deforming objects where traditional visual hull intersection approaches fail to reconstruct the object correctly.

25. Video Falsifying by Motion Interpolation and Inpainting

Timothy Shih, Nick Tang, Tsai Joseph, Zhong Hsing-Ying

We change the behavior of actors in a video. For instance, the outcome of a 100-meter race in the Olympic game can be falsified. We track objects and segment motions using a modified mean shift mechanism. The resulting video layers can be played in different speeds and at different reference points with respect to the original video. In order to obtain a smooth movement of target objects, a motion interpolation mechanism is proposed based on continuous stick figures (i.e., a video of human skeleton) and video inpainting. The video inpainting mechanism is performed in a quasi-3D space via guided 3D patch matching for filling. Interpolated target objects and background layers are fused by using graph cut. It is hard to tell whether a falsified video is the original. We demonstrate the original and the falsified videos in our website at http://www.mine.tku.edu.tw/video_demo/). The proposed technique can be used to create special effects in movie industry.

26. Dynamic scene shape reconstruction using a single structured light pattern

Hiroshi Kawasaki, Ryo Furukawa, Ryusuke Sagawa, Yasushi Yagi

3D acquisition techniques to measure dynamic scenes and deformable objects with little texture are extensively researched for applications like the motion capturing of human facial expression. To allow such measurement, several techniques using structured light have been proposed. These techniques can be largely categorized into two types. The first involves techniques to temporally encode positional information of a projector's pixels using multiple projected patterns, and the second involves techniques to spatially encode positional information into areas or color spaces. Although the former allows dense reconstruction with a sufficient number of patterns, it has difficulty in scanning objects in rapid motion. The latter technique uses only a single pattern, so this problem can be resolved, however, it often uses complex patterns or color intensities, which are weak to noise, shape distortions, or textures. Thus, it remains an open problem to achieve dense and stable 3D acquisition in real cases. In this paper, we propose a technique to achieve dense shape reconstruction that requires only a single-frame image of a grid pattern. The proposed technique also has the advantage of being robust in terms of image processing.

27. Simultaneous super-resolution and 3D video using graph-cuts

Tony Tung, Shohei Nobuhara, Takashi Matsuyama

This paper presents a new method to increase the quality of 3D video, a new media developed to represent 3D objects in motion. This representation is obtained from multi-view reconstruction techniques that require images recorded simultaneously by several video cameras. All cameras are calibrated and placed around a dedicated studio to fully surround the models. The limited quality and quantity of cameras may produce inaccurate 3D model reconstruction with low quality texture. To overcome this issue, first we propose super-resolution (SR) techniques for 3D video: SR on multi-view images and SR on single-view video frames. Second, we propose to combine both super-resolution and dynamic 3D shape reconstruction problems into a unique Markov Random Field (MRF) energy formulation. The MRF minimization is performed using graph-cuts. Thus, we jointly compute the optimal solution for super-resolved texture and 3D shape model reconstruction. Moreover, we propose a coarse-to-fine strategy to iteratively produce 3D video with increasing quality. Our experiments show the accuracy and robustness of the proposed technique on challenging 3D video sequences.

28. Stereoscopic Inpainting: Joint Color and Depth Completion from Stereo Images

Liang Wang, Hailin Jin, Ruigang Yang, Minglun Gong

We present a novel algorithm for simultaneous color and depth inpainting. The algorithm takes stereo images and estimated disparity maps as input and fills in missing color and depth information introduced by occlusions or object removal. We first complete the disparities for the occlusion regions using a segmentation-based approach. The completed disparities can be used to facilitate the user in labeling objects to be removed. Since part of the removed regions in one image is visible in the other, we mutually complete the two images through 3D warping. Finally, we complete the remaining unknown regions using a depth-assisted texture synthesis technique, which simultaneously fills in both color and depth. We demonstrate the effectiveness of the proposed algorithm on several challenging data sets.

29. Retinal Image Registration from 2D to 3D

Yuping Lin, Gérard Medioni

We propose a 2D registration method for multi-modal image sequences of the retinal fundus, and a 3D metric reconstruction of near planar surface from multiple views. There are two major contributions in our paper. For 2D registration, our method produces high registration rates while accounting for large modality differences. Compared with the state of the art method, our approach has higher registration rate (97.2% vs. 82.31%) while the computation time is much less. This is achieved by extracting features from the edge maps of the contrast enhanced images, and performing pairwise registration by matching the features in an iterative manner, maximizing the number of matches and estimating homographies accurately. The pairwise registration result is further globally optimized by an indirect registration process. For 3D registration part, images are registered to the reference frame by transforming points via a reconstructed 3D surface. The challenge is the reconstruction of a near planar surface, in which the shallow depth makes it a quasi-degenerate case for estimating the geometry from images. Our contribution is the proposed 4-pass bundle adjustment method that gives optimal estimation of all camera poses. With accurate camera poses, the 3D surface can be reconstructed using the images associated with the cameras with the largest baseline. Compared with state of the art 3D retinal image registration methods, our approach produces better results in all image sets.

30. On Benchmarking Camera Calibration And Multi-View Stereo

Christoph Strecha, Wolfgang von Hansen, Luc Van Gool, Pascal Fua, Ulrich Thoennessen

In this paper we want to start the discussion on whether image based 3-D modelling techniques can possibly be used to replace LIDAR systems for outdoor 3D data acquisition. Two main issues have to be addressed in this context: (*i*) camera calibration (internal and external) and (*ii*) dense multi-view stereo. To investigate both, we have acquired test data from outdoor scenes both with LIDAR and cameras. Using the LIDAR data as reference we estimated the ground-truth for several scenes. Evaluation sets are prepared to evaluate different aspects of 3D model building. These are: (*i*) pose estimation and multi-view stereo with known internal camera parameters; (*ii*) camera calibration and multi-view stereo with the raw images as the only input and (*iii*) multi-view stereo.

31. What Can Missing Correspondences Tell Us About 3D Structure and Motion?

Christopher Zach, Arnold Irschara, Horst Bischof

Practically all existing approaches to structure and motion computation use only positive image correspondences to verify the camera pose hypotheses. Incorrect epipolar geometries are solely detected by identifying outliers among the found correspondences. Ambigous patterns in the images are often incorrectly handled by these standard methods. In this work we propose two approaches to overcome such problems. First, we apply non-monotone reasoning on view triplets using a Bayesian formulation. In contrast to two-view epipolar geometry, image triplets allow the prediction of features in the third image. Absence of these features (i.e., missing correspondences) enables additional inference about the view triplet. Furthermore, we integrate these view triplet handling into an incremental procedure for structure and motion computation. Thus, our approach is able to refine the maintained 3D structure when additional image data is provided.

32. A Factorization Approach to Structure from Motion with Shape Priors

Alessio Del Bue

This paper presents an approach for including 3D prior models into a factorization framework for structure from motion. The proposed method computes a closed-form affine fit which mixes the information from the data and the 3D prior on the shape structure. Moreover, it is general in regards to different classes of objects treated: rigid, articulated and deformable. The inclusion of the shape prior may aid the inference of camera motion and 3D structure components whenever the data is degenerate (i.e., nearly planar motion of the projected shape). A final non-linear optimization stage, which includes the shape priors as a quadratic cost, upgrades the affine fit to metric. Results on real and synthetic image sequences, which present predominant degenerate motion, make clear the improvements over the 3D reconstruction.

33. Photogeometric Structured Light: A Self-Calibrating and Multi-Viewpoint Framework for Accurate 3D Modeling

Daniel G. Aliaga, Yi Xu

Structured-light methods actively generate geometric correspondence data between projectors and cameras in order to facilitate robust 3D reconstruction. In this paper, we present Photogeometric Structured Light whereby a standard structured light method is extended to include photometric methods. Photometric processing serves the double purpose of increasing the amount of recovered surface detail and of enabling the structured-light setup to be robustly self-calibrated. Further, our framework uses a photogeometric optimization that supports the simultaneous use of multiple cameras and projectors and yields a single and accurate multi-view 3D model which best complies with photometric and geometric data.

34. Minimal Solutions for Generic Imaging Models

Srikumar Ramalingam, Peter Sturm

A generic imaging model refers to a non-parametric camera model where every camera is treated as a set of unconstrained projection rays. Calibration would simply be a method to map the projection rays to image pixels; such a mapping can be computed using plane based calibration grids. However, existing algorithms for generic calibration use more point correspondences than the theoretical minimum. It has been well-established that non-minimal solutions for calibration and structure-from-motion algorithms are generally noise-prone compared to minimal solutions. In this work we derive minimal solutions for generic calibration algorithms. Our algorithms for generally central cameras use 4 point correspondences in three calibration grids to compute the motion between the grids. Using simulations we show that our minimal solutions are more robust to noise compared to non-minimal solutions. We also show very accurate distortion correction results on fisheye images.

35. Real-Time Global Localization with A Pre-Built Visual Landmark Database

Zhiwei Zhu, Taragay Oskiper, Supun Samarasekera, Rakesh Kumar, Harpreet S. Sawhney

In this paper, we study how to build a vision-based system for global localization with accuracies within 10 cm for robots and humans operating both indoors and outdoors over wide areas covering many square kilometers. In particular, we study the parameters of building a landmark database rapidly and utilizing that database online for real-time accurate global localization. Although the accuracy of traditional short-term motion based visual odometry systems has improved significantly in recent years, these systems alone cannot solve the drift problem over large areas. Landmark based localization combined with visual odometry is a viable solution to the large scale localization problem. However, a systematic study of the specification and use of such a landmark database has not been undertaken.

We propose techniques to build and optimize a landmark database systematically and efficiently using visual odometry. First, topology inference is utilized to find overlapping images in the database. Second, bundle adjustment is used to refine the accuracy of each 3D landmark. Finally, the database is optimized to balance the size of the database with achievable accuracy. Once the landmark database is obtained, a new real-time global localization methodology that works both indoors and outdoors is proposed. We present results of our study on both synthetic and real datasets that help us determine critical design parameters for the landmark database and the achievable accuracies of our proposed system.

36. Photometric Stereo with Coherent Outlier Handling and Confidence Estimation

Frank Verbiest, Luc Van Gool

In photometric stereo a robust method is required to deal with outliers, such as shadows and non-Lambertian reflections. In this paper we rely on a probabilistic imaging model that distinguishes between inliers and outliers, and formulate the problem as a Maximum-Likelihood estimation problem. To signal which imaging model to use a hidden binary inlier map is introduced, which, to account for the fact that inlier/outlier pixels typically group together, is modelled as a Markov Random Field. To make inference of model parameters and hidden variables tractable a mean field Expectation-Maximization (EM) algorithm is used. If for each pixel we add the scaled normal, i.e., albedo and normal combined, to the model parameters, it would not be possible to obtain a confidence estimate in the result. Instead, each scaled normal is added as a hidden variable, the distribution of which, approximated by a Gaussian, is also estimated in the EM algorithm. The covariance matrix of the recovered approximate Gaussian distribution serves as a confidence estimate of the scaled normal. We demonstrate experimentally the effectiveness or our approach.

37. Fast Algorithms for L¥ Problems in Multiview Geometry

Sameer Agarwal, Noah Snavely, Steven Seitz

Many problems in multi-view geometry, when posed as minimization of the maximum reprojection error across observations, can be solved optimally in polynomial time. We show that these problems are instances of a convex-concave generalized fractional program. We survey the major solution methods for solving problems of this form and present them in a unified framework centered around a single parametric optimization problem. We propose two new algorithms and show that the algorithm proposed by Olsson et al. is a special case of a classical algorithm for generalized fractional programming. The performance of all the algorithms is compared on a variety of datasets, and the algorithm proposed by Gugat stands out as a clear winner. An open source MATLAB toolbox thats implements all the algorithms presented here is made available.

38. Robust Null Space Representation and Sampling for View-Invariant Motion Trajectory Analysis

Xu Chen, Dan Schonfeld, Ashfaq Khokhar

In this paper, we propose a novel robust retrieval and classification system for video and motion events based on null space representation. In order to analyze the robustness of the system, the perturbed null operators have been derived with the first order perturbation theory. Subsequently, the sensitivity of the null operators is discussed in terms of the error ratio and the SNR respectively. Meanwhile, the normwise bounds and componentwise bounds based on classical matrix perturbation theory are presented and discussed. Given the perturbation, uniform sampling are proposed for the convergence of the SNR and Poisson sampling are proposed for the convergence of the error ratio in the mean sense by choosing the rate parameter the same order as the number of samples. The simulation results are provided to demonstrate the effectiveness and robustness of our system in motion event indexing, retrieval and classification that is invariant to affine transformation due to camera motions.

39. Spatio-temporal Saliency Detection Using the Phase Spectrum of Quaternion Fourier Transform

Chenlei Guo, Qi Ma, Liming Zhang

Salient areas in natural scenes are generally regarded as the candidates of attention focus in human eyes, which is the key stage in object detection. In computer vision, many models have been proposed to simulate the behavior of eyes such as SaliencyToolBox (STB), Neuromorphic Vision Toolkit (NVT) and etc., but they demand high computational cost and their remarkable results mostly rely on the choice of parameters. Recently a simple and fast approach based on Fourier transform called spectral residual (SR) was proposed, which used SR of the amplitude spectrum to obtain the saliency map. The results are good, but the reason is questionable.

In this paper, we propose it is the phase spectrum, not the amplitude spectrum, of the Fourier transform that is the key in obtaining the location of salient areas. We provide some examples to show that PFT can get better results in comparison with SR and requires less computational complexity as well. Furthermore, PFT can be easily extended from a two-dimensional Fourier transform to a Quaternion Fourier Transform (QFT) if the value of each pixel is represented as a quaternion composed of intensity, color and motion feature. The added motion dimension allows the phase spectrum to represent spatio-temporal saliency in order to engage in attention selection for videos as well as images.

Extensive tests of videos, natural images and psychological patterns show that the proposed method is more effective than other models. Moreover, it is very robust against white-colored noise and meets the real-time requirements, which has great potentials in engineering applications.

40. View and Scale Invariant Action Recognition Using Multiview Shape-Flow Models

Pradeep Natarajan, Ram Nevatia

Actions in real world applications typically take place in cluttered environments with large variations in the orientation and scale of the actor. We present an approach to simultaneously track and recognize known actions that are robust to such variations, starting from a person detection in the standing pose. In our approach we first render synthetic poses from multiple viewpoints using Mocap data for known actions and represent them in a Conditional Random Field (CRF) whose observation potentials are computed using shape similarity and the transition potentials are computed using optical flow. We enhance these basic potentials with terms to represent spatial and temporal constraints and call our enhanced model the *Shape, Flow, Duration-Conditional Random Field(SFD-CRF)*. We find the best sequence of actions using Viterbi search in the SFD-CRF. We demonstrate our approach on videos from multiple viewpoints and in the presence of background clutter.

41. Facial Expression Recognition Based on Dynamic Binary Patterns

Peng Yang, Liu Qingshan, Cui Xinyi, Dimitris Metaxas

In this paper, we propose a novel framework for video-based facial expression recognition, which can handle the data with various time resolutions including a single frame. We first use the haar-like features to represent facial appearance, due to their simplicity and effectiveness. Then we perform *K*-Means clustering on the facial appearance features to explore the intrinsic temporal patterns of each expression. Based on the temporal pattern models, we further map the facial appearance variations into dynamic binary patterns. Finally, boosting learning is performed to construct the expression classifiers. Compared to previous work, the dynamic binary patterns encode the intrinsic dynamics of expression, and our method makes no assumption on the time resolution of the data. Extensive experiments carried on the Cohn-Kanade database show the promising performance of the proposed method.

42. Trajectory Analysis and Semantic Region Modeling Using A Nonparametric Bayesian Model

Xiaogang Wang, Eric Grimson

We propose a novel nonparametric Bayesian model, Dual Hierarchical Dirichlet Processes (Dual-HDP), for trajectory analysis and semantic region modeling in surveillance settings, in an unsupervised way. In our approach, trajectories are treated as documents and observations of an object on a trajectory are treated as words in a document. Trajectories are clustered into different activities. Abnormal trajectories are detected as samples with low likelihoods. The semantic regions, which are intersections of paths commonly taken by objects, related to activities in the scene are also modeled. Dual-HDP advances the existing Hierarchical Dirichlet Processes (HDP) language model. HDP only clusters co-occurring words from documents into topics and automatically decides the number of topics. Dual-HDP co-clusters both words and documents. It learns both the numbers of word topics and document clusters from data. Under our problem settings, HDP only clusters observations of objects, while Dual-HDP clusters both observations and trajectories. Experiments are evaluated on two data sets, radar tracks collected from a maritime port and visual tracks collected from a parking lot.

43. Manifold-Manifold Distance with Application to Face Recognition based on Image Set

Ruiping Wang, Shiguang Shan, Xilin Chen, Wen Gao

In this paper, we address the problem of classifying image sets, each of which contains images belonging to the same class but covering large variations in, for instance, viewpoint and illumination. We innovatively formulate the problem as the computation of Manifold-Manifold Distance (MMD), i.e., calculating the distance between nonlinear manifolds each representing one image set. To compute MMD, we also propose a novel manifold learning approach, which expresses a manifold by a collection of local linear models, each depicted by a subspace. MMD is then converted to integrating the distances between pair of subspaces respectively from one of the involved manifolds. The proposed MMD method is evaluated on the task of Face Recognition based on Image Set (FRIS). In FRIS, each known subject is enrolled with a set of facial images and modeled as a gallery manifold, while a testing subject is modeled as a probe manifold, which is then matched against all the gallery manifolds by MMD. Identification is achieved by seeking the minimum MMD. Experimental results on two public face databases, Honda/UCSD and CMU MoBo, demonstrate that the proposed MMD method outperforms the competing methods.

44. Near Duplicate Image Identification with Spatially Aligned Pyramid Matching

Dong Xu, Tat Jen Cham, Shuicheng Yan, Shih-Fu Chang

A new framework, termed Spatially Aligned Pyramid Matching, is proposed for Near Duplicate Image Identification. The proposed method robustly handles spatial shifts as well as scale changes. Images are divided into both overlapped and non-overlapped blocks over multiple levels. In the first matching stage, pairwise distances between blocks from the examined image pair are computed using SIFT features and Earth Mover's Distance (EMD). In the second stage, multiple alignment hypotheses that consider piecewise spatial shifts and scale variation are postulated and resolved using integer-flow EMD. Two application scenarios are addressed -- retrieval ranking and binary classification. For retrieval ranking, a pyramid-based scheme is constructed to fuse matching results from different partition levels. For binary classification, a novel Generalized Neighborhood Component Analysis method is formulated that can be effectively used in tandem with SVMs to select the most critical matching components. The proposed methods are shown to clearly outperform existing methods through extensive testing on the Columbia Near Duplicate Image Database and another new dataset.

45. Pose primitive based human action recognition in videos or still images

Christian Thurau, Václav Hlaváč

This paper presents a method for recognizing human actions based on pose primitives. In learning mode, the parameters representing poses and activities are estimated from videos. In run mode, the method can be used both for videos or still images. For recognizing pose primitives, we extend a Histogram of Oriented Gradient (HOG) based descriptor to better cope with articulated poses and cluttered background. Action classes are represented by histograms of poses primitives. For sequences, we incorporate the local temporal context by means of n-gram expressions. Action recognition is based on a simple histogram comparison. Unlike the mainstream video surveillance approaches, the proposed method does not rely on background subtraction or dynamic features and thus allows for action recognition in still images.

46. Correspondence-Free Multi-Camera Activity Analysis and Scene Modeling

Xiaogang Wang, Kinh Tieu, Eric Grimson

We propose a novel approach for activity analysis in multiple synchronized but uncalibrated static camera views. We assume that the topology of camera views is unknown and quite arbitrary, the fields of views covered by these cameras may have no overlap or any amount of overlap, and objects may move on different ground planes. Using low-level cues, objects are tracked in each of the camera views independently, and the positions and velocities of objects along trajectories are computed as features. Under a generative model, our approach jointly learns the distribution of an activity in the feature spaces of different camera views. It accomplishes two tasks: (1) grouping trajectories in different camera views belonging to the same activity into one cluster; (2) modeling paths commonly taken by objects across camera views. To our knowledge, no prior result of co-clustering trajectories in multiple camera views has been published. Advantages of this approach are that it does not require first solving the challenging correspondence problem, and the learning is unsupervised. Our approach is evaluated on two very large data sets with 22, 951 and 14, 985 trajectories.

47. Learning Human Actions via Information Maximization

Liu Jingen, Shah Mubarak

In this paper, we present a novel approach for automatically learning a compact and yet discriminative appearance-based human action model. A video sequence is represented by a bag of spatiotemporal features called video-words by quantizing the extracted 3D interest points (cuboids) from the videos. Our proposed approach is able to automatically discover the optimal number of video-word clusters by utilizing Maximization of Mutual Information(MMI). Unlike the *k*-means algorithm, which is typically used to cluster spatiotemporal cuboids into video words based on their appearance similarity, MMI clustering further groups the *video-words*, which are highly correlated to some group of actions. To capture the structural information of the learnt optimal video-word clusters, we explore the correlation of the compact video-word clusters. We use the modified correlgoram, which is not only translation and rotation invariant, but also somewhat scale invariant. We extensively test our proposed approach on two publicly available challenging datasets: the KTH dataset and IXMAS multiview dataset. To the best of our knowledge, we are the first to try the bag of video-words related approach on the multiview dataset. We have obtained very impressive results on both datasets.

48. Learning Human Motion Models from Unsegmented Videos

Roman Filipovych, Eraldo Ribeiro

We present a novel method for learning human motion models from unsegmented videos. We propose a unified framework that encodes spatio-temporal relationships between descriptive motion parts and the appearance of individual poses. Sparse sets of spatial and spatio-temporal features are used. The method automatically learns static pose models and spatio-temporal motion parts. Neither motion cycles nor human figures need to be segmented for learning. We test the model on a publicly available action dataset and demonstrate that our new method performs well on a number of classification tasks. We also show that classification rates are improved by increasing the number of pose models in the framework.

49. Unsupervised Learning of Human Perspective Context Using ME-DT for Efficient Human Detection in Surveillance

Liyuan Li, Maylor K. H. Leung

A novel and automated technique for learning human perspective context (HPC) from a scene is proposed in this paper. It is found that two models are required to describe HPC for camera tilt angle ranging from 0° to 50°. From a scene, the tilt angle can be inferred from the observed human shapes and head/foot positions. Afterward, a novel ME-DT (Model Estimation - Data Tuning) algorithm is proposed to learn human perspective context from live data of various degrees of uncertainties. The uncertainties may come from the variations of human individual heights and poses, and segmentation/recognition errors. ME-DT not only estimates the model parameters from the training data but also tunes the data to achieve a better head-foot correlation. The human perspective context provides a feasible constraint on the scales, positions, and orientations of humans in the scene. Applying this constraint to the HOG human detection, great reduction of the detection windows and improved performances have been obtained compared to conventional methods.

50. Recognizing Primitive Interactions by Exploring Actor-Object States

Roman Filipovych, Eraldo Ribeiro

In this paper, we present a solution to the novel problem of recognizing primitive actor-object interactions from videos. Here, we introduce the concept of actor-object states. Our method is based on the observation that at the moment of physical contact, both the motion and the appearance of actors are constrained by the target object. We propose a probabilistic framework that automatically learns models in such constrained states. We use joint probability distributions to represent both actor and object appearances as well as their intrinsic spatio-temporal configurations. Finally, we demonstrate the applicability of our approach on series of human-object interaction classification experiments.

51. Action MACH: A spatio-temporal maximum average correlation height filter for action recognition

Mikel Rodriguez, Javed Ahmed, Shah Mubarak

In this paper we introduce a template-based method for recognizing human actions called Action MACH. Our approach is based on a Maximum Average Correlation Height (MACH) filter. A common limitation of template-based methods is their inability to generate a single template using a collection of examples. MACH is capable of capturing intra-class variability by synthesizing a single Action MACH filter for a given action class. We generalize the traditional MACH filter to video (3D spatiotemporal volume), and vector valued data. By analyzing the response of the filter in the frequency domain, we avoid the high computational cost commonly incurred in template-based approaches. Vector valued data is analyzed using the Clifford Fourier transform, a generalization of the Fourier transform intended for both scalar and vector-valued data. Finally, we perform an extensive set of experiments and compare our method with some of the most recent approaches in the field by using publicly available datasets, and two new annotated human action datasets which include actions performed in classic feature films and sports broadcast television.

52. Simultaneous Clustering and Tracking Unknown Number of Objects

Katsuhiko Ishiguro, Takeshi Yamada, Naonori Ueda

In this paper, we present a novel on-line probabilistic generative model that simultaneously deals with both the clustering and the tracking of an unknown number of moving objects. The proposed model assumes that i) time series data are composed of a time-varying number of objects and that ii) each object is governed by a mixture of an unknown number of different patterns of dynamics. The problem of learning patterns of dynamics is formulated as the clustering of tracked objects based on a nonparametric Bayesian model with conjugate priors, and this clustering in turn improves the tracking. We present a particle filter for posterior estimation of simultaneous clustering and tracking. Through experiments with synthetic and real movie data, we confirmed that the proposed model successfully learned the hidden cluster patterns and obtained better tracking results than conventional models without clustering.

53. Pan, Zoom, Scan - Time-coherent, Trained Automatic Video Cropping

Thomas Deselaers, Philippe Dreuw, Hermann Ney

We present a method to fully automatically fit videos in 16:9 format on 4:3 screens and vice versa. It can be applied to arbitrary aspect ratios and can be used to make videos suitable for mobile viewing devices with small and possibly uncommonly sized displays. The cropping sequence is optimised over time to create smooth transitions and thus leads to an excellent viewing experience. Current televisions have simple and often disturbing methods which either show the centre region of the image, distort the image, or pad it with black borders. The technique presented here can fully automatically find the “right” viewing area for each image in a video sequence. It works in real-time with only very little time-shift. We employ different low-level features and a log-linear model to learn how to find the right area. The method is able to automatically decide whether padding with black borders is necessary or whether all relevant image areas fit on screen by cropping the image. Evaluation is done on ten videos from five different types of content and the baseline methods are clearly outperformed.

54. Action Snippets: How many frames does human action recognition require?

Konrad Schindler, Luc Van Gool

Visual recognition of human actions in video clips has been an active field of research in recent years. However, most published methods either analyse an entire video and assign it a single action label, or use relatively large look-ahead to classify each frame. Contrary to these strategies, human vision proves that simple actions can be recognised almost instantaneously. In this paper, we present a system for action recognition from very short sequences (“snippets”) of 1--10 frames, and systematically evaluate it on standard data sets. It turns out that even local shape and optic flow for a single frame are enough to achieve ~90% correct recognitions, and snippets of 5-7 frames (0.3-0.5 seconds of video) are enough to achieve a performance similar to the one obtainable with the entire video sequence.

55. Action Recognition using Exemplar-based Embedding

Daniel Weinland, Edmond Boyer

In this paper, we address the problem of representing human actions using visual cues for the purpose of learning and recognition. Traditional approaches model actions as space-time representations which explicitly or implicitly encode the dynamics of an action through temporal dependencies. In contrast, we propose a new compact and efficient representation which does not account for such dependencies. Instead, motion sequences are represented with respect to a set of discriminative static key-pose exemplars and without modeling any temporal ordering. The interest is a time-invariant representation that drastically simplifies learning and recognition by removing time related information such as speed or length of an action. The proposed representation is equivalent to embedding actions into a space defined by distances to key-pose exemplars. We show how to build such embedding spaces of low dimension by identifying a vocabulary of highly discriminative exemplars using a forward selection. To test our representation, we have used a publicly available dataset which demonstrates that our method can precisely recognize actions, even with cluttered and non-segmented sequences.

56. Human Action Recognition using Local Spatio-Temporal Discriminant Embedding

Kui Jia, Dit-Yan Yeung

Human action video sequences can be considered as nonlinear dynamic shape manifolds in the space of image frames. In this paper, we address learning and classifying human actions on embedded low-dimensional manifolds. We propose a novel manifold embedding method, called Local Spatio-Temporal Discriminant Embedding (LSTDE). The discriminating capabilities of the proposed method are two-fold: (1) for local spatial discrimination, LSTDE projects data points (silhouette-based image frames of human action sequences) in a local neighborhood into the embedding space where data points of the same action class are close while those of different classes are far apart; (2) in such a local neighborhood, each data point has an associated short video segment, which forms a local temporal subspace on the embedded manifold. LSTDE finds an optimal embedding which maximizes the principal angles between those temporal subspaces associated with data points of different classes. Benefiting from the joint spatio-temporal discriminant embedding, our method is potentially more powerful for classifying human actions with similar space-time shapes, and is able to perform recognition on a frame-by-frame or short video segment basis. Experimental results demonstrate that our method can accurately recognize human actions, and can improve the recognition performance over some representative manifold embedding methods, especially on highly confusing human action types.

57. Statistical analysis on Stiefel and Grassmann Manifolds with applications in Computer Vision

Pavan Turaga, Ashok Veeraraghavan, Rama Chellappa

Many applications in computer vision and pattern recognition involve drawing inferences on certain manifold-valued parameters. In order to develop accurate inference algorithms on these manifolds we need to a) understand the geometric structure of these manifolds b) derive appropriate distance measures and c) develop probability distribution functions (pdf) and estimation techniques that are consistent with the geometric structure of these manifolds. In this paper, we consider two related manifolds - the Stiefel manifold and the Grassmann manifold, which arise naturally in several vision applications such as spatio-temporal modeling, affine invariant shape analysis, image matching and learning theory. We show how accurate statistical characterization that reflects the geometry of these manifolds allows us to design efficient algorithms that compare favorably to the state of the art in these very different applications. In particular, we describe appropriate distance measures and parametric and non-parametric density estimators on these manifolds. These methods are then used to learn class conditional densities for applications such as activity recognition, video based face recognition and shape classification.

58. Macro-cuboids based probabilistic matching for lip-reading digits

Samuel Pachoud, Shaogang Gong, Andrea Cavallaro

In this paper, we present a spatio-temporal feature representation and a probabilistic matching function to recognise lip movements from pronounced digits. Our model (1) automatically selects spatio-temporal features extracted from 10 digit model templates and (2) matches them with probe video sequences. Spatio-temporal features embed lip movements from pronouncing digits and contain more discriminative information than spatial features alone. A model template for each digit is represented by a set of spatio-temporal features at multiple scales. A probabilistic sequence matching function automatically segments a probe video sequence and matches the most likely sequence of digits recognised in the probe sequence. We demonstrate the proposed approach using the CUAVE database and compare our representational scheme with three alternative methods, based on optical flow, intensity gradient and block matching, respectively. The evaluation shows that the proposed approach outperforms the others in recognition accuracy and is robust in coping with variations in probe sequences.

59. Action Recognition by Learning Mid-Level Motion Features

Alireza Fathi, Greg Mori

This paper presents a method for human action recognition based on patterns of motion. Previous approaches to action recognition use either local features describing small patches or large-scale features describing the entire human figure. We develop a method constructing mid-level motion features which are built from low-level optical flow information. These features are focused on local regions of the image sequence and are created using a variant of AdaBoost. These features are tuned to discriminate between different classes of action, and are efficient to compute at run-time. A battery of classifiers based on these mid-level features is created and used to classify input sequences. State-of-the-art results are presented on a variety of standard datasets.

60. Tensor Reduction Error Analysis -- Applications to Video Compression and Classification

Chris Ding, Heng Huang, Dijun Luo

Tensor based dimensionality reduction has recently been extensively studied for computer vision applications. To our knowledge, however, there exist no rigorous error analysis on these methods. Here we provide the first error analysis of these methods and provide error bound results similar to Eckart-Young Theorem which plays critical role in the development and application of singular value decomposition (SVD). Beside performance guarantee, these error bounds are useful for subspace size determination according to the required video/image reconstruction error. Furthermore, video surveillance/retrieval, 3D/4D medical image analysis, and other computer vision applications require particular reduction in spatio-temporal space, but not along data index dimension. This motivates a *D*-1 tensor reduction. Standard method such as high order SVD (HOSVD) compress data in all index dimensions and thus can not perform the classification and pattern recognition tasks. We provide algorithm and error bound analysis of the *D*-1 factorization for spatio-temporal data dimensionality. Experiments on video sequences demonstrate our approach outperforms the previous dimensionality deduction methods for spatio-temporal data.

61. Learning 4D Action Feature Models for Arbitrary View Action Recognition

Pingkun Yan, Saad Khan, Mubarak Shah

In this paper we present a novel approach using a 4D (x,y,z,t) action feature model (4D-AFM) for recognizing actions from arbitrary views. The 4D-AFM elegantly encodes shape and motion of actors observed from multiple views. The modeling process starts with reconstructing 3D visual hulls of actors at each time instant. Spatiotemporal action features are then computed in each view by analyzing the differential geometric properties of spatio-temporal volumes (3D STVs) generated by concatenating the actor's silhouette over the course of the action (x,y,t). These features are mapped to the sequence of 3D visual hulls over time (4D) to build the initial 4D-AFM. Actions are recognized based on the scores of matching action features from the input videos to the model points of 4D-AFMs by exploiting pairwise interactions of features. Promising recognition results have been demonstrated on the multi-view IXMAS dataset using both single and multi-view input videos.

62. Detection with Multi-exit Asymmetric Boosting

Minh-Tri Pham, Viet-Dung D. Hoang, Tat-Jen Cham

We introduce a generalized representation for a boosted classifier with multiple exit nodes, and propose a method to training which combines the idea of propagating scores across boosted classifiers and the use of asymmetric goals. A means for determining the ideal constant asymmetric goal is provided, which is theoretically justified under a conservative bound on the ROC operating point target and empirically near-optimal under the exact bound. Moreover, our method automatically minimizes the number of weak classifiers, avoiding the need to retrain a boosted classifier multiple times for empirical best performance as in conventional methods. Experimental results shows significant reduction in training time and number of weak classifiers, as well as better accuracy, compared to conventional cascades and multi-exit boosted classifiers.

63. Multiplicative Kernels: Object Detection, Segmentation and Pose Estimation

Quan Yuan, Ashwin Thangali, Vitaly Ablavsky, Stan Sclaroff

Object detection is challenging when the object class exhibits large within-class variations. In this work, we show that foreground-background classification (detection) and within-class classification of the foreground class (pose estimation) can be jointly learned in a multiplicative form of two kernel functions. One kernel measures similarity for foreground-background classification. The other kernel accounts for latent factors that control within-class variation and implicitly enables feature sharing among foreground training samples. Detector training can be accomplished via standard SVM learning. The resulting detectors are tuned to specific variations in the foreground class. They also serve to evaluate hypotheses of the foreground state. When the foreground parameters are provided in training, the detectors can also produce parameter estimate. When the foreground object masks are provided in training, the detectors can also produce object segmentation. The advantages of our method over past methods are demonstrated on data sets of human hands and vehicles.

64. Simultaneous Data Volume Reconstruction and Pose Estimation from Slice Samples

Manfred Georg, Richard Souvenir, Andrew Hope, Robert Pless

Modeling the dynamics of heart and lung tissue is challenging because the tissue deforms between data acquisitions. To reconstruct complete volumes, sample data captured at different times and locations must be combined. This paper presents a novel end-to-end, data driven framework for the complete reconstruction of deforming tissue volumes. This framework is a joint optimization over an undeformed tissue volume, a deformation map that describes tissue motion for given pose parameters (i.e., breathing and heartbeat), and an estimate of those parameters for each data acquisition. Tissue motion is modeled by deforming a reference volume with a cubic B-spline free form deformation, and we use Isomap to derive initial estimates of the pose of sample data. An iterative method is used to simultaneously solve for the reference volume and deformation map while updating the pose estimates. This same process is demonstrated on 4D CT lung data and heart/lung MR data.

65. Efficient subdivision-based image and volume warping

Gady Agam, Ravinder Singh

Warping is fundamental to multiple algorithms in computer vision and medical imaging such as image and volume registration. Warping is performed by determining a continuous deformation map and applying it to a given image or volume. In registration the deformation map is determined based on correspondence between two images. It is often the case that the deformation map can only be determined at discrete locations and so has to be interpolated. The discrete locations where the deformation map is determined form irregular sampling of the unknown continuous deformation map. Thin-plate splines are commonly used to perform the interpolation and provide an optimal solution in the sense of bending energy minimization. Assuming N samples of the deformation map and n^2 image pixels, thin plate splines require solving a $N \times N$ dense linear system with $O(N^3)$ complexity for determining spline coefficients and N computations per pixel with $O(N\,n^2)$ complexity for determining interpolated values. When N and n are large as in the case of volumetric medical image analysis this cost becomes prohibitive. The approach proposed in this paper is based on subdivision surfaces and is capable of achieving similar quality results with $O(N \log N)$ complexity for coefficient determination and $O(n^2)$ complexity for computing interpolated values. Experimental results demonstrate two orders of magnitude performance improvement on actual clinical data.

66. Volumetric Reconstruction from Multi-Energy Single-View Radiography

Sang N. Le, Mei Kay Lee, Shamima Banu, Anthony C. Fang

We address the volumetric reconstruction problem that takes as input a series of orthographic multi-energy x-ray images, producing as output a reconstructed model space consisting of uniform-size mass density voxels. Our approach solves the non-linear constrained optimization formulation problem by constructing a compliant estimate of volumetric distribution, subject to projective and domain constraints, and minimizes variational irregularities. To resolve the inherent ambiguities of single-view formulation, an optional shape model may be introduced to aid the reconstruction process. We demonstrate our method's practical usage as a new in-vivo method for estimating three-dimensional body segmental compositions, and compare its results with the contemporary methods.

67. Probabilistic Image Registration and Anomaly Detection by Nonlinear Warping

Verena Kaynig, Bernd Fischer, Joachim M. Buhmann

Automatic, defect tolerant registration of transmission electron microscopy (TEM) images poses an important and challenging problem for biomedical image analysis, e.g., in computational neuroanatomy. In this paper we demonstrate a fully automatic stitching and distortion correction method for TEM images and propose a probabilistic approach for image registration. The technique identifies image defects due to sample preparation and image acquisition by outlier detection. A polynomial kernel expansion is used to estimate a non-linear image transformation based on intensities and spatial features. Corresponding points in the images are not determined beforehand, but they are estimated via an EM-algorithm during the registration process which is preferable in the case of (noisy) TEM images. Our registration model is successfully applied to two large image stacks of serial section TEM images acquired from brain tissue samples in a computational neuroanatomy project and shows significant improvement over existing image registration methods on these large datasets.

68. Global Image Registration Based on Learning the Prior Appearance Model

Ayman El-Baz, Georgy Gimel`farb

A new approach to align an image of a textured object with a given prototype (learned reference object) is proposed. Visual appearance of the images, after equalizing their signals, is modeled with a Markov-Gibbs random field with pairwise interaction. Similarity to the prototype (learned reference object) is measured by a Gibbs energy of signal co-occurrences in a characteristic subset of pixel pairs derived automatically from the prototype. An object is aligned by an affine transformation maximizing the similarity by using an automatic initialization followed by gradient search. To get accurate appearance model, we developed a new approach to automatically select the most important cliques (neighborhood system) that describe the visual appearance of a texture object. Experiments confirm that our approach aligns complex objects better than popular conventional algorithms.

69. Probabilistic multi-tensor estimation using the tensor distribution function

Alex Leow, Siwei Zhu, Katie McMahon, Greig I. de Zubicaray, Matt Meredith, Margie Wright, Paul Thompson

Diffusion weighted magnetic resonance (MR) imaging is a powerful tool that can be employed to study white matter microstructure by examining the 3D displacement profile of water molecules in brain tissue. By applying diffusion-sensitized gradients along a minimum of 6 directions, second-order tensors can be computed to model dominant diffusion processes. However, conventional DTI is not sufficient to resolve crossing fiber tracts. Recently, a number of high-angular resolution schemes with greater than 6 gradient directions have been employed to address this issue. In this paper, we introduce the Tensor Distribution Function (TDF), a probability function defined on the space of symmetric positive definite matrices. Here, fiber crossing is modeled as an ensemble of Gaussian diffusion processes with weights specified by the TDF. Once this optimal TDF is determined, the diffusion orientation distribution function (ODF) can easily be computed by analytic integration of the resulting displacement probability function.

70. Active Microscopic Cellular Image Annotation by Superposable Graph Transduction with Imbalanced Labels

Jun Wang, Shih-Fu Chang, Xiaobo Zhou, Stephen T. C. Wong

Systematic content screening of cell phenotypes in microscopic images has been shown promising in gene function understanding and drug design. However, manual annotation of cells and images in genome-wide studies is cost prohibitive. In this paper, we propose a highly efficient active annotation framework, in which a small amount of expert input is leveraged to rapidly and effectively infer the labels over the remaining unlabeled data. We formulate this as a graph based transductive learning problem and develop a novel method for label propagation. Specifically, a label regularizer method is proposed to handle the important label imbalance issue, typically seen in the cellular image screening applications. We also design a new scheme which breaks the graph into linear superposition of contributions from individual labeled samples. We take advantage of such a superposable representation to achieve fast annotation in an interactive setting. Extensive evaluations over toy data and realistic cellular images confirm the superiority of the proposed method over existing alternatives.

71. Robust Statistics on Riemannian Manifolds via the Geometric Median

P. Thomas Fletcher, Suresh Venkatasubramanian, Sarang Joshi

The geometric median is a classic robust estimator of centrality for data in Euclidean spaces. In this paper we formulate the geometric median of data on a Riemannian manifold as the minimizer of the sum of geodesic distances to the data points. We prove existence and uniqueness of the geometric median on manifolds with non-positive sectional curvature and give sufficient conditions for uniqueness on positively curved manifolds. Generalizing the Weiszfeld procedure for finding the geometric median of Euclidean data, we present an algorithm for computing the geometric median on an arbitrary manifold. We show that this algorithm converges to the unique solution when it exists. This method produces a robust central point for data lying on a manifold, and should have use in a variety of vision applications involving manifolds. We give examples of the geometric median computation and demonstrate its robustness for three types of manifold data: the 3D rotation group, tensor manifolds, and shape spaces.

72. Adaptive Region Intensity Based Rigid Ultrasound and CT Image Registration

Zhijun Zhang

Rigid registration of intraoperative ultrasound (US) and CT is an important technique to provide real-time guidance for preoperative images and models. Due to the speckle noise and artefacts in US images, accurate registration of CT and US is still a challenging problem. We propose an adaptive region intensity based CT and US registration method. The registration is initialized by matching the distinctive regions of CT and US images. Then the registration is a multistage process in which the regions in US used will be adaptively updated at each stage. The registration problem is considered as a global similarity energy optimization and high local statistical dependency regions selection process. Performances of our method and other intensity based method are evaluated with simulated and real datasets. Experiments results show the improvement of our registration method in robustness and accuracy.

73. Optimizing Discrimination-Efficiency Tradeoff in Integrating Heterogeneous Local Features for Object Detection

Bo Wu, Ram Nevatia

A large variety of image features has been invented for detection of objects of a known class. We propose a framework to optimize the discrimination-efficiency tradeoff in integrating multiple, heterogeneous features for object detection. Cascade structured detectors are learned by boosting local feature based weak classifiers. Each weak classifier corresponds to a local image region, from which several different types of features are extracted. The weak classifier makes predictions by examining the features one by one; this classifier goes to the next feature only when the prediction from the already examined features is not confident enough. The order in which the features are evaluated is determined based on their computational cost normalized classification powers. We apply our approach to two object classes, pedestrians and cars. The experimental results show that our approach outperforms the state-of-the-art methods.

74. Segmentation of Multiple, Partially Occluded Objects by Grouping, Merging, Assigning Part Detection Responses

Bo Wu, Ram Nevatia, Yuan Li

We propose a method that detects and segments multiple, partially occluded objects in images. A part hierarchy is defined for the object class. Whole-object segmentor and part detectors are learned by boosting shape oriented local image features. During detection, the part detectors are applied to the input image. All the edge pixels in the image that positively contribute to part detection responses are extracted. A joint likelihood of multiple objects is defined based on the part detection responses and the object edges. Computing the joint likelihood includes an inter-object occlusion reasoning that is based on the object silhouettes extracted with the whole-object segmentor. By maximizing the joint likelihood, part detection responses are grouped, merged, and assigned to multiple object hypotheses. The proposed approach is applied to the pedestrian class, and evaluated on two public test sets. The experimental results show that our method outperforms the previous ones.

75. FlowFusion: Discrete-Continuous Optimization for Optical Flow Estimation

Victor Lempitsky, Stefan Roth, Carsten Rother

Accurate estimation of optical flow is a challenging task, which often requires addressing difficult energy optimization problems. To solve them, most top-performing methods rely on continuous optimization algorithms. The modeling accuracy of the energy in this case is often traded for its tractability. This is in contrast to the related problem of narrow-baseline stereo matching, where the top-performing methods employ powerful discrete optimization algorithms such as graph cuts and message-passing to optimize highly non-convex energies.

In this paper, we demonstrate how similar non-convex energies can be formulated and optimized discretely in the context of optical flow estimation. Starting with a set of candidate solutions that are produced by fast continuous flow estimation algorithms, the proposed method iteratively fuses these candidate solutions by the computation of minimum cuts on graphs. The obtained continuous-valued fusion result is then further improved using local gradient descent. Experimentally, we demonstrate that the proposed energy is an accurate model and that the proposed discrete-continuous optimization scheme not only finds lower energy solutions than traditional discrete or continuous optimization techniques, but also leads to flow estimates that outperform the current state-of-the-art.

10:30am – 12:15pm Oral Session O3A-1: Face, Gesture, and Action (Cook)

1. Model-Based Hand tracking with texture, Shading and Self-occlusions

Martin de La Gorce, Nikos Paragios, David J. Fleet

A novel model-based approach to 3D hand tracking from monocular video is presented. The 3D hand pose, the hand texture and the illuminant are dynamically estimated through minimization of an objective function. Derived from an inverse problem formulation, the objective function enables explicit use of texture temporal continuity and shading information, while handling important self-occlusions and time-varying illumination. The minimization is done efficiently using a quasi-Newton method, for which we propose a rigorous derivation of the objective function gradient. Particular attention is given to terms related to the change of visibility near self-occlusion boundaries that are neglected in existing formulations. In doing so we introduce new occlusion forces and show that using all gradient terms greatly improves the performance of the method. Experimental results demonstrate the potential of the formulation.

2. Face Alignment via Boosted Ranking Model

Hao Wu, Xiaoming Liu, Gianfranco Doretto

Face alignment seeks to deform a face model to match it with the features of the image of a face by optimizing an appropriate cost function. We propose a new face model that is aligned by maximizing a score function, which we learn from training data, and that we impose to be concave. We show that this problem can be reduced to learning a classifier that is able to say whether or not by switching from one alignment to a new one, the model is approaching the correct fitting. This relates to the ranking problem where a number of instances need to be ordered. For training the model, we propose to extend GentleBoost to rank-learning. Extensive experimentation shows the superiority of this approach to other learning paradigms, and demonstrates that this model exceeds the alignment performance of the state-of-the-art.

3. Learning Patch Correspondences for Improved Viewpoint Invariant Face Recognition

Ahmed Bilal Ashraf, Simon Lucey, Tsuhan Chen

Variation due to viewpoint is one of the key challenges that stand in the way of a complete solution to the face recognition problem. It is easy to note that local regions of the face change differently in appearance as the viewpoint varies. Recently, patch-based approaches, such as those of Kanade and Yamada, have taken advantage of this effect resulting in improved viewpoint invariant face recognition. In this paper we propose a data-driven extension to their approach, in which we not only model how a face patch varies in appearance, but also how it deforms spatially as the viewpoint varies. We propose a novel alignment strategy which we refer to as “stack flow” that discovers viewpoint induced spatial deformities undergone by a face at the patch level. One can then view the spatial deformation of a patch as the correspondence of that patch between two viewpoints. We present improved identification and verification results to demonstrate the utility of our technique.

4. View-Invariant Action Recognition Using Fundamental Ratios

Yuping Shen, Hassan Foroosh

A moving plane observed by a fixed camera induces a fundamental matrix **F** across multiple frames, where the ratios among the elements in the upper left 2×2 submatrix are herein referred to as the *Fundamental Ratios*. We show that fundamental ratios are invariant to camera parameters, and hence can be used to identify similar plane motions from varying viewpoints. For action recognition, we decompose a body posture into a set of point triplets (planes). The similarity between two actions is then determined by the motion of point triplets and hence by their associated fundamental ratios, providing thus view-invariant recognition of actions. Results evaluated over 255 semi-synthetic video data with 100 independent trials at a wide range of noise levels, and also on 56 real videos of 8 different classes of actions, confirm that our method can recognize actions under substantial amount of noise, even when they have dynamic timeline maps, and the viewpoints and camera parameters are unknown and totally different.

5. Learning realistic human actions from movies

Ivan Laptev, Marcin Marszałek, Cordelia Schmid, Benjamin Rozenfeld

The aim of this paper is to address recognition of natural human actions in diverse and realistic video settings. This challenging but important subject has mostly been ignored in the past due to several problems one of which is the lack of realistic and annotated video datasets. Our first contribution is to address this limitation and to investigate the use of movie scripts for automatic annotation of human actions in videos. We evaluate alternative methods for action retrieval from scripts and show benefits of a text-based classifier. Using the retrieved action samples for visual learning, we next turn to the problem of action classification in video. We present a new method for video classification that builds upon and extends several recent ideas including local space-time features, space-time pyramids and multi-channel non-linear SVMs. The method is shown to improve state-of-the-art results on the standard KTH action dataset by achieving 91.8% accuracy. Given the inherent problem of noisy labels in automatic annotation, we particularly investigate and show high tolerance of our method to annotation errors in the training set. We finally apply the method to learning and classifying challenging action classes in movies and show promising results.

10:30am – 12:15pm Oral Session O3A-2: Correspondence and Registration (La Perouse)

1. A Polynomial-Time Bound for Matching and Registration with Outliers

Carl Olsson, Olof Enqvist, Fredrik Kahl

We present a framework for computing optimal transformations, aligning one point set to another, in the presence of outliers. Example applications include shape matching and registration (using, for example, similarity, affine or projective transformations) as well as multiview reconstruction problems (triangulation, camera pose etc.).

While standard methods like RANSAC essentially use heuristics to cope with outliers, we seek to find the largest possible subset of consistent correspondences and the globally optimal transformation aligning the point sets. Based on theory from computational geometry, we show that this is indeed possible to accomplish in polynomial-time. We develop several algorithms which make efficient use of convex programming. The scheme has been tested and evaluated on both synthetic and real data for several applications.

2. Dense Correspondence Finding for Parametrization-free Animation Reconstruction from Video

Naveed Ahmed, Christian Theobalt, Christian Röessl, Sebastian Thrun, Hans-Peter Seidel

We present a dense 3D correspondence finding method that enables spatio-temporally coherent reconstruction of surface animations from multi-view video data. Given as input a sequence of shape-from-silhouette volumes of a moving subject that were reconstructed for each time frame individually, our method establishes dense surface correspondences between subsequent shapes independently of surface discretization. This is achieved in two steps: first, we obtain sparse correspondences from robust optical features between adjacent frames. Second, we generate dense correspondences which serve as map between respective surfaces. By applying this procedure subsequently to all pairs of time steps we can trivially align one shape with all others. Thus, the original input can be reconstructed as a sequence of meshes with constant connectivity and small tangential distortion. We exemplify the performance and accuracy of our method using several synthetic and captured real-world sequences.

3. Matching images under unstable segmentations

Varsha Hedau, Himanshu Arora, Narendra Ahuja

Region based features are getting popular due to their higher descriptive power relative to other features. However, real world images exhibit changes in image segments capturing the same scene part taken at different time, under different lighting conditions, from different viewpoints, etc. Segmentation algorithms reflect these changes, and thus segmentations exhibit poor repeatability. In this paper we address the problem of matching regions of similar objects under unstable segmentations. Merging and splitting of regions makes it difficult to find such correspondences using one-to-one matching algorithms. We present partial region matching as a solution to this problem. We assume that the high contrast, dominant contours of an object are fairly repeatable, and use them to compute partial matching cost (PMC) between regions. Region correspondences are obtained under region adjacency constraints encoded by Region Adjacency Graph (RAG). We integrate PMC in a many-to-one label assignment framework for matching RAGs, and solve it using belief propagation. We show that our algorithm can match images of similar objects across unstable image segmentations. We also compare the performance of our algorithm with that of the standard one-to-one matching algorithm on three motion sequences. We conclude that our partial region matching approach is robust under segmentation irrepeatabilities.

4. Adaptive Parametrization of Multivariate B-splines for Image Registration

Michael Sass Hansen, Ben Glocker, Nassir Navab, Rasmus Larsen

We present an adaptive parametrization scheme for dynamic mesh refinement in the application of parametric image registration. The scheme is based on a refinement measure ensuring that the control points give an efficient representation of the warp fields, in terms of minimizing the registration cost function. In the current work we introduce multivariate B-splines as a novel alternative to the widely used tensor B-splines enabling us to make efficient use of the derived measure.

The multivariate B-splines of order n are C^{n-1} smooth and are based on Delaunay configurations of arbitrary 2D or 3D control point sets. Efficient algorithms for finding the configurations are presented, and B-splines are through their flexibility shown to feature several advantages over the tensor B-splines. In spite of efforts to make the tensor product B-splines more flexible, the knots are still bound to reside on a regular grid. In contrast, by efficient non-constrained placement of the knots, the multivariate B-splines are shown to give a good representation of inhomogeneous objects in natural settings.

The wide applicability of the method is illustrated through its application on medical data and for optical flow estimation.

5. Fusion of Time-of-Flight Depth and Stereo for High Accuracy Depth Maps

Jiejie Zhu, Liang Wang, Ruigang Yang, James Davis

Time-of-flight range sensors have error characteristics which are complementary to passive stereo. They provide real time depth estimates in conditions where passive stereo does not work well, such as on white walls. In contrast, these sensors are noisy and often perform poorly on the textured scenes for which stereo excels. We introduce a method for combining the results from both methods that performs better than either alone. A depth probability distribution function from each method is calculated and then merged. In addition, stereo methods have long used global methods such as belief propagation and graph cuts to improve results, and we apply these methods to this sensor. Since time-of-flight devices have primarily been used as individual sensors, they are typically poorly calibrated. We introduce a method that substantially improves upon the manufacturer's calibration. We show that these techniques lead to improved accuracy and robustness.

1:45pm – 3:45pm Poster Session P3P-1: Selected Topics (Summit Hall)

1. Recursive photometric stereo when multiple shadows and highlights are present

Vasileios Argyriou, Maria Petrou

We present a recursive algorithm for 3D surface reconstruction based on Photometric Stereo in the presence of highlights, and self and cast shadows. We assume that the surface reflectance outside the highlights can be approximated by the Lambertian model. The algorithm works with as few as three light sources, and it can be generalised for N without any difficulties. Furthermore, this reconstruction method is able to identify areas where the majority of the lighting directions result in unreliable pixel intensities, providing the capability to adjust a reconstruction algorithm and improve its performance avoiding the unreliable sources. We report results for both artificial and real images and compare them with the results of other state of the art photometric stereo algorithms.

2. On controlling light transport in poor visibility environments

Mohit Gupta, Srinivasa G. Narasimhan, Yoav Y. Schechner

Poor visibility conditions due to murky water, bad weather, dust and smoke severely impede the performance of vision systems. Passive methods have been used to restore scene contrast under moderate visibility by digital post-processing. However, these methods are ineffective when the quality of acquired images is poor to begin with. In this work, we design active lighting and sensing systems for controlling light transport before image formation, and hence obtain higher quality data. First, we present a technique of polarized light striping based on combining polarization imaging and structured light striping. We show that this technique out-performs different existing illumination and sensing methodologies. Second, we present a numerical approach for computing the optimal relative sensor-source position, which results in the best quality image. Our analysis accounts for the limits imposed by sensor noise.

3. A Two-Frame Theory of Motion, Lighting and Shape

Ronen Basri, Darya Frolova

This paper explores how shape, motion, and lighting interact in the case of a two-frame motion sequence. We consider a rigid object with Lambertian reflectance properties undergoing small motion with respect to both a camera and a stationary point light source. Assuming orthographic projection, we derive a single, first order quasilinear partial differential equation that relates shape, motion, and lighting, while eliminating out the albedo. We show how this equation can be solved, when the motion and lighting parameters are known, to produce a 3D reconstruction of the object. A solution is obtained using the method of characteristics and can be refined by adding regularization. We further show that both smooth bounding contours as well as surface markings can be used to derive Dirichlet boundary conditions. Experimental results demonstrate the quality of this reconstruction.

4. Bayesian Color Constancy Revisited

Peter Gehler, Carsten Rother, Toby Sharp, Andrew Blake, Tom Minka

Computational color constancy is the task of estimating the true reflectances of visible surfaces in an image. In this paper we follow a line of research that assumes uniform illumination of a scene, and that the principal step in estimating reflectances is the estimation of the scene illuminant. We review recent approaches to illuminant estimation, firstly those based on formulae for normalisation of the reflectance distribution in an image -- so-called grey-world algorithms, and those based on a Bayesian formulation of image formation. In evaluating these previous approaches we introduce a new tool in the form of a database of 568 high-quality, indoor and outdoor images, accurately labelled with illuminant, and preserved in their raw form, free of correction or normalisation. This has enabled us to establish several properties experimentally. Firstly automatic selection of grey-world algorithms according to image properties is not nearly so effective as has been thought. Secondly, it is shown that Bayesian illuminant estimation is significantly improved by the improved accuracy of priors for illuminant and reflectance that are obtained from the new dataset.

5. An LED-only BRDF Measurement Device

Moshe Ben-Ezra, Jiaping Wang, Bennett Wilburn, Xiaoyang Li, Le Ma

Light Emitting Diodes (LEDs) can be used as light detectors and as light emitters. In this paper, we present a novel BRDF measurement device consisting exclusively of LEDs. Our design can acquire BRDFs over a full hemisphere, or even a full sphere (for the bidirectional transmittance distribution function BTDF) , and can also measure a (partial) multi-spectral BRDF. Because we use no cameras, projectors, or even mirrors, our design does not suffer from occlusion problems. It is fast, significantly simpler, and more compact than existing BRDF measurement designs.

6. Characterizing the Shadow Space of Camera-Light Pairs

Daniel A. Vaquero, Rogerio S. Feris, Matthew Turk, Ramesh Raskar

We present a theoretical analysis for characterizing the shadows cast by a point light source given its relative position to the camera. In particular, we analyze the epipolar geometry of camera-light pairs, including unusual camera-light configurations such as light sources aligned with the camera's optical axis as well as convenient arrangements such as lights placed in the camera plane. A mathematical characterization of the shadows is derived to determine the orientations and locations of depth discontinuities when projected onto the image plane that could potentially be associated with cast shadows. The resulting theory is applied to compute a lower bound on the number of lights needed to extract all depth discontinuities from a general scene using a multiflash camera. We also provide a characterization of which discontinuities are missed and which are correctly detected by the algorithm, and a foundation for choosing an optimal light placement. Experiments with depth edges computed using two-flash setups and a four-flash setup illustrate the theory, and an additional configuration with a flash at the camera's center of projection is exploited as a solution for some degenerate cases.

7. Overcoming Visual Reverberations

Yaron Diamant, Yoav Y. Schechner

An image acquired through a glass window is a superposition of two sources: a scene behind the window, and a reflection of a scene in front of the window. Light rays incident on the window are reflected back and forth inside the glass. Such internal reflections affect the radiance of both sources: a spatial effect is created of dimmed and shifted replications. Our work generalizes the treatment of transparent scenes to deal with this effect. First, we present a physical model of the image formation. It turns out that each of the transmitted and reflected scenes undergoes a convolution with a particular point spread function (PSF), composed of distinct delta functions. Therefore, scene recovery involves inversion of these PSFs. We analyze the fundamental limitations faced by any attempt to solve this inverse problem. We then present a solution approach. The approach is based on deconvolution by linear filtering and simple optimization. The input to the algorithm is a pair of frames, taken through a polarizing filter. The method is demonstrated experimentally.

8. 3D Shape Reconstruction of Mooney Faces

Ira Kemelmacher-Shlizerman, Ronen Basri, Boaz Nadler

Two-tone (“Mooney”) images seem to arouse vivid 3D percept of faces, both familiar and unfamiliar, despite their seemingly poor content. Recent psychological and fMRI studies suggest that this percept is guided primarily by top-down procedures in which recognition precedes reconstruction. In this paper we investigate this hypothesis from a mathematical standpoint. We show that indeed, under standard shape from shading assumptions, a Mooney image can give rise to multiple different 3D reconstructions even if reconstruction is restricted to the Mooney transition curve (the boundary curve between black and white) alone. We then construct top-down methods for 3D shape reconstruction of novel faces from single Mooney images exploiting prior knowledge of the structure of at least one face of a different individual. We apply the methods to thresholded images of real faces and demonstrate the reconstruction quality relative to reconstruction from gray level images.

9. Beyond the Lambertian Assumption: A generative model for Apparent BRDF field for Faces using Anti-Symmetric Tensor Splines

Angelos Barmpoutis, Ritwik Kumar, Baba C. Vemuri, Arunava Banerjee

Human faces are neither exactly Lambertian nor entirely convex and hence most models in literature which make the Lambertian assumption, fall short when dealing with specularities and cast shadows. In this paper, we present a novel anti-symmetric tensor spline (a spline for tensor-valued functions) based method for the estimation of the Apparent BRDF (ABRDF) field for human faces that seamlessly accounts for specularities and cast shadows. Furthermore, unlike other methods, it does not require any 3D information to build the model and can work with as few as 9 images. In order to validate the accuracy of our anti-symmetric tensor spline model, we present a novel approximation of the ABRDF using a continuous mixture of single-lobed spherical functions. We demonstrate the effectiveness of our anti-symmetric tensor-spline model in comparison to other popular models in literature, by presenting extensive results for face relighting and face recognition using the Extended Yale B database.

10. Efficient Photometric Stereo on Glossy Surfaces with Wide Specular Lobes

Hin Shun Chung, Jiaya Jia

This paper presents a new photometric stereo method aiming to efficiently estimate BRDF and reconstruct glossy surfaces. Rough specular surfaces exhibit wide specular lobes under different lightings. They are ubiquitous and usually bring difficulties to both specular pixel removal and surface normal recovery. In our approach, we do not apply unreliable highlight separation and specularity estimation. Instead, an important visual cue, i.e., the cast shadow silhouette of the object, is employed to optimally recover global BRDF parameters. These parameter estimates are then taken into a reflectance model for robustly computing the surface normals and other local parameters using an iterative optimization. Within the unified framework, our method can also be used to reconstruct object surfaces assembled with multiple materials.

11. A NURBS-Based Spectral Reflectance Descriptor with Applications in Computer Vision and Pattern Recognition

Cong Phuoc Huynh, Antonio Robles-Kelly

In this paper, we present a surface reflectance descriptor based on the control points resulting from the interpolation of Non-Uniform Rational B-Spline (NURBS) curves to multispectral reflectance data. The interpolation is based upon a knot removal scheme in the parameter domain. Thus, we exploit the local support of NURBS so as to recover a compact descriptor robust to noise and local perturbation of the spectra. We demonstrate the utility of our NURBSbased descriptor for material identification. To this end, we perform skin spectra recognition making use of a Support Vector Machine classifier. We also provide results on hyperspectral imagery and elaborate on the preprocessing step for skin segmentation. We compare our results with those obtained using an alternative descriptor.

12. Shading Models for Illumination and Reflectance Invariant Shape Detectors

Peter Nillius, Josephine Sullivan, Antonis Argyros

Many objects have smooth surfaces of a fairly uniform color, thereby exhibiting shading patterns that reveal information about its shape, an important clue to the nature of the object. This papers explores extracting this information from images, by creating shape detectors based on shading.

Recent work has derived low-dimensional models of shading that can handle realistic unknown lighting conditions and surface reflectance properties. We extend this theory by also incorporating variations in the surface shape. In doing so it enables the creation of very general models for the 2D appearance of objects, not only coping with variations in illumination and BRDF but also in shape alterations such as small scale and pose changes. Using this framework we propose a scheme to build shading models that can be used for shape detection in a bottom up fashion without any a priori knowledge about the scene.

From the developed theory we construct detectors for two basic shape primitives, spheres and cylinders. Their performance is evaluated by extensive synthetic experiments as well as experiments on real images.

13. Recovering Shape Characteristics on Near-flat Specular Surfaces

Yuanyuan Ding, Jingyi Yu

We consider the problem of capturing shape characteristics on specular (refractive and reflective) surfaces that are nearly flat. These surfaces are difficult to model using traditional methods based on reconstructing the surface positions and normals. These lower-order shape attributes provide little information to identify important surface characteristics related to distortions. In this paper, we present a framework for recovering the higher-order geometry attributes of specular surfaces. Our method models local reflections and refractions in terms of a special class of multiperspective cameras called the general linear cameras (GLCs). We then develop a new theory that correlates the higher-order differential geometry attributes with the local GLCs. Specifically, we show that Gaussian and Mean Curvature can be directly derived from the camera intrinsics of the local GLCs. We validate this theory on both synthetic and real-world specular surfaces. Our method places a known pattern in front of a reflective surface or beneath a refractive surface and captures a distorted image on the surface. We then compute the optimal GLC using a sparse set of correspondences and recover the curvatures from the GLC. Experiments demonstrate that our methods are robust and highly accurate.

14. Efficient Object Shape Recovery via Slicing Planes

Po-Lun Lai, Alper Yilmaz

Recovering the three-dimensional (3D) object shape remains an unresolved area of research on the cross-section of computer vision, photogrammetry and bioinformatics. Although various techniques have been developed, the computational complexity and the constraints introduced to overcome the problems have limited their applicability in the real world scenarios. In this paper, we propose a method that is based on the projective geometry between the object space and the silhouette-images taken from multiple viewpoints. The approach eliminates the problems related to dense feature point matching and camera calibration that are generally adopted by many state of the art shape reconstruction methods. The object shape is reconstructed by establishing a set of hypothetical planes slicing the object volume and estimating the projective geometric relations between the images of these planes. The experimental results show that the 3D object shape can be recovered by applying minimal constraints.

15. 3D Surface Models by Geometric Constraints Propagation

Michela Farenzena, Andrea Fusiello

This paper proposes a technique for estimating piecewise planar models of objects from their images and geometric constraints. First, assuming a bounded noise in the localization of 2D points, the position of the 3D point is estimated as a polyhedron containing all the possible solutions of the triangulation. Then, given the topological structure of the 3D points cloud, geometric relationships among facets, such as coplanarity, parallelism, orthogonality, and angle equality, are automatically detected. A subset of them that is sufficient to stabilize the 3D model estimation is selected with a flow-network based algorithm. Finally a feasible instance of the 3D model, i.e., one that satisfies the selected geometric relationships and whose 3D points lie within the associated polyhedral bounds, is computed by solving a Constraint Satisfaction Problem.

16. Toward Automatic 3D Modeling of Scenes using a Generic Camera Model

Maxime Lhuillier

The automatic reconstruction of 3D models from image sequences is still a very active field of research. All existing methods are designed for a given camera model, and a new (and ambitious) challenge is 3D modeling with a method which is exploitable for any kind of camera. A similar approach was recently suggested for structure-from-motion thanks to the use of generic camera models. In this paper, we first introduce geometric tools designed for 3D scene modeling with a generic camera model. Then, these tools are used to solve many issues: matching errors, wide range of point depths, depth discontinuities, and view-point selection for reconstruction. Experiments are provided for perspective and catadioptric cameras.

17. Building reconstruction from a single DEM

Florent Lafarge, Xavier Descombes, Josiane Zerubia, Marc Pierrot-Deseilligny

We present a new approach for building reconstruction from a single Digital Elevation Model (DEM). It treats buildings as an assemblage of simple urban structures extracted from a library of 3D parametric blocks (like a LEGO® set). This method works on various data resolutions such as 0.7 m satellite and 0.1 m aerial DEMs and allows us to obtain 3D representations with various levels of detail. First, the 2D supports of the urban structures are extracted either interactively or automatically. Then, 3D blocks are placed on the 2D supports using a Gibbs model. A Bayesian decision finds the optimal configuration of 3D blocks using a RJMCMC sampler. Experimental results on complex buildings and dense urban areas are presented using data at various resolutions.

18. Recovery of relative depth from a single observation using an uncalibrated (real-aperture) camera

Vinay P. Namboodiri, Subhasis Chaudhuri

In this paper we investigate the challenging problem of recovering the depth layers in a scene from a single defocused observation. The problem is definitely solvable if there are multiple observations. In this paper we show that one can perceive the depth in the scene even from a single observation. We use the inhomogeneous reverse heat equation to obtain an estimate of the blur, thereby preserving the depth information characterized by the defocus. However, the reverse heat equation, due to its parabolic nature, is divergent. We stabilize the reverse heat equation by considering the gradient degeneration as an effective stopping criterion. The amount of (inverse) diffusion is actually a measure of relative depth. Because of ill-posedness we propose a graph-cuts based method for inferring the depth in the scene using the amount of diffusion as a data likelihood and a smoothness condition on the depth in the scene. The method is verified experimentally on a varied set of test cases.

19. Practical Camera Auto-Calibration based on Object Appearance and Motion for Traffic Scene Visual Surveillance

Zhaoxiang Zhang, Min Li, Kaiqi Huang, Tieniu Tan

Camera calibration, as a fundamental issue in computer vision, is indispensable in many visual surveillance applications. Firstly, calibrated camera can help to deal with perspective distortion of object appearance on image plane. Secondly, calibrated camera makes it possible to recover metrics from images which are robust to scene or view angle changes. In addition, with calibrated cameras, we can make use of prior information of 3D models to estimate 3D pose of objects and make object detection or tracking more robust to noise and occlusions.

In this paper, we propose an automatic method to recover camera models from traffic scene surveillance videos. With only the camera height H measured, we can completely recover both intrinsic and extrinsic parameters of cameras based on appearance and motion of objects in videos. Experiments are conducted in different scenes and experimental results demonstrate the effectiveness and practicability of our approach, which can be adopted in many traffic scene surveillance applications.

20. Joint Data Alignment Up To (Lossy) Transformations

Andrea Vedaldi, Gregorio Guidi, Stefano Soatto

Joint data alignment is often regarded as a data simplification process. This idea is powerful and general, but raises two delicate issues. First, one must make sure that the useful information about the data is preserved by the alignment process. This is especially important when data are affected by non-invertible transformations, such as those originating from continuous domain deformations in a discrete image lattice. We propose a formulation that explicitly avoids this pitfall. Second, one must choose an appropriate measure of data complexity. We show that standard concepts such as entropy might not be optimal for the task, and we propose alternative measures that reflect the regularity of the codebook space. We also propose a novel and efficient algorithm that allows joint alignment of a large number of samples (tens of thousands of image patches), and does not rely on the assumption that pixels are independent. This is done for the case where the data is postulated to live in an affine subspaces of the embedding space of the raw data. We apply our scheme to learn sparse bases for natural images that discount domain deformations and hence significantly decrease the complexity of codebooks while maintaining the same generative power.

21. High Quality Mesostructure Acquisition Using Specularities

Yannick Francken, Tom Cuypers, Tom Mertens, Jo Gielis, Philippe Bekaert

We propose a technique for cheap and efficient acquisition of mesostructure normal maps from specularities, which only requires a simple LCD monitor and a digital camera. Coded illumination enables us to capture subtle surface details with only a handful of images. In addition, our method can deal with heterogeneous surfaces, and high albedo materials. We are able to recover highly detailed mesostructures, which was previously only possible with an expensive hardware setup.

22. Image Based Rendering for Motion Compensation in Angiographic Roadmapping

Christian Unger, Martin Groher, Nassir Navab

2D angiographic roadmapping is used frequently during image guided interventions to superimpose vessel structures onto currently acquired fluoroscopic images. While the fluoroscopic images, acquired with 12-15 frames per second, show patient bone anatomy as well as the current location of the inserted catheter, the roadmap delineates vessels to provide path information and to avoid accidental vessel wall punctures during catheter advancement.

This technique successfully reduces the injection of contrast agent, which is hazardous for the patient; however, it suffers from inaccuracy due to inevitable patient movement, which yields a misalignment of the roadmap laid over the current fluoroscopic frame.

We propose a method for rigid patient motion compensation via the trifocal tensor and Image Based Rendering (IBR). The method uses two contrasted and slightly shifted views and the current fluoroscopic frame.

Different to the existing solutions, we perform the motion compensation inherently in 3D, increasing reliability and accuracy of the resulting vascular rendering. Moreover, with the IBR technique, we avoid an explicit reconstruction, thus achieving reasonable results even for very small patient movements, which are common in interventional scenarios.

23. IM2GPS: estimating geographic information from a single image

James Hays, Alexei A. Efros

Estimating geographic information from an image is an excellent, difficult high-level computer vision problem whose time has come. The emergence of vast amounts of geographically-calibrated image data is a great reason for computer vision to start looking globally -- on the scale of the entire planet! In this paper, we propose a simple algorithm for estimating a distribution over geographic locations from a single image using a purely data-driven scene matching approach. For this task, we will leverage a dataset of over 6 million GPS-tagged images from the Internet. We represent the estimated image location as a probability distribution over the Earth's surface. We quantitatively evaluate our approach in several geolocation tasks and demonstrate encouraging performance (up to 30 times better than chance). We show that geolocation estimates can provide the basis for numerous other image understanding tasks such as population density estimation, land cover estimation or urban/rural classification.

24. Sketching in the Air: A Vision-Based System for 3D Object Design

Yu Chen, Jianzhuang Liu, Xiaoou Tang

3D object design has many applications including flexible 3D sketch input in CAD, computer game, webpage content design, image based object modeling, and 3D object retrieval. Most current 3D object design tools work on a 2D drawing plane such as a computer screen or tablet, which is often inflexible with one dimension lost. On the other hand, virtual reality based methods have the drawbacks that there are awkward devices worn by the user and the virtual environment systems are expensive. In this paper, we propose a novel vision-based approach to 3D object design. Our system consists of a PC, a camera, and a mirror. We use the camera and mirror to track a wand so that the user can design 3D objects by sketching in 3D free space directly without having to wear any cumbersome devices. A number of new techniques are developed for working in this system, including input of object wireframes, gestures for editing and drawing objects, and optimization-based planar and curved surface generation. Our system provides designers a new user interface for designing 3D objects conveniently.

25. Multi-Object Shape Estimation and Tracking from Silhouette Cues

Li Guan, Jean-Sebastien Franco, Marc Pollefeys

This paper deals with the 3D shape estimation from silhouette cues of multiple moving objects in general indoor or outdoor 3D scenes with potential static obstacles, using multiple calibrated video streams. Most shape-from-silhouette techniques use a two-classification of space occupancy and silhouettes, based on image regions that match or disagree with a static background appearance model. Binary silhouette information becomes insufficient to unambiguously carve 3D space regions as the number and density of dynamic objects increases. In such difficult scenes, multi-view stereo methods suffer from visibility problems, and rely on color calibration procedures tedious to achieve outdoors. We propose a new algorithm to automatically detect and reconstruct scenes with a variable number of dynamic objects. Our formulation distinguishes between m different shapes in the scene by using automatically learned view-specific object appearance models, eliminating the color calibration requirement. Bayesian reasoning is then applied to solve the m-shape occupancy problem, with m updated as objects enter or leave the scene. Results show that this method yields multiple silhouette-based estimates that drastically improve scene reconstructions over traditional two-label silhouette scene analysis. This enables the method to also efficiently deal with multi-person tracking problems.

26. Max Margin AND/OR Graph Learning for Parsing the Human Body

Long Zhu, Yuanhao Chen, Yifei Lu, Chenxi Lin, Alan Yuille

We present a novel structure learning method, Max Margin AND/OR Graph (MM-AOG), for parsing the human body into parts and recovering their poses. Our method represents the human body and its parts by an AND/OR graph, which is a multi-level mixture of Markov Random Fields (MRFs). Max-margin learning, which is a generalization of the training algorithm for support vector machines (SVMs), is used to learn the parameters of the AND/OR graph model discriminatively. There are four advantages from this combination of AND/OR graphs and max-margin learning. Firstly, the AND/OR graph allows us to handle enormous articulated poses with a compact graphical model. Secondly, max-margin learning has more discriminative power than the traditional maximum likelihood approach. Thirdly, the parameters of the AND/OR graph model are optimized globally. In particular, the weights of the appearance model for individual nodes and the relative importance of spatial relationships between nodes are learnt simultaneously. Finally, the kernel trick can be used to handle high dimensional features and to enable complex similarity measure of shapes. We perform comparison experiments on the baseball datasets, showing significant improvements over state of the art methods.

27. Automatic Calibration of a Single-Projector Catadioptric Display System

Benjamin Astre, Laurent Sarry, Christophe Lohou, Eric Zeghers

We describe the calibration of a catadioptric omnidirectional video-projection system that adjusts its projection to the geometry of any scene by means of a rotating camera. Correction of geometric distortions requires 3D reconstruction of the scene. A camera is used to detect projected point features and calibration is performed in three successive steps: precalibration of camera assuming pure rotation, precalibration of catadioptric projector under central approximation, and calibration of the global system, by minimizing the squared distance between the reflected and perceived rays, and by relaxing previous constraints, to refine values of extrinsic parameters. Simulation is used to validate estimated values of parameters and distance between the 3D reconstruction of the projection room and its expected geometry. Influence of noise in detected point coordinates is studied and preliminary results for the reconstruction and projection in real conditions are reported.

28. SMRFI: Shape Matching via Registration of Vector-Valued Feature Images

Lisa Tang, Ghassan Hamarneh

We perform shape matching by transforming the problem of establishing shape correspondences into an image registration problem. At each vertex on the shape, we calculate a shape feature and encode this feature as image intensity at appropriate positions in the image domain. Calculating multiple features at each vertex and encoding them into the image domain results in a vector-valued feature image. Establishing point correspondence between two shapes is thereafter treated as a registration problem of two vector-valued feature images. With this shape representation, various existing image registration strategies can now be easily applied. These include the use of a scale-space approach to diffuse the shape features, a coarse-to-fine registration scheme, and various deformable registration algorithms. As our validation shows, by representing shapes as vector-valued images, the overall method is robust against noise and occlusions. To this end, we have successfully established 2D point correspondences of shapes of corpora callosa, vertebrae, and brain ventricles.

29. From Skeletons to Bone Graphs: Medial Abstraction for Object Recognition

Diego Macrini, Kaleem Siddiqi, Sven Dickinson

Medial descriptions, such as shock graphs, have gained significant momentum in the shape-based object recognition community due to their invariance to translation, rotation, scale and articulation and their ability to cope with moderate amounts of within-class deformation. While they attempt to decompose a shape into a set of parts, this decomposition can suffer from ligature-induced instability. In particular, the addition of even a small part can have a dramatic impact on the representation in the vicinity of its attachment. We present an algorithm for identifying and representing the ligature structure, and restoring the non-ligature structures that remain. This leads to a bone graph, a new medial shape abstraction that captures a more intuitive notion of an object's parts than a skeleton or a shock graph, and offers improved stability and within-class deformation invariance. We demonstrate these advantages by comparing the use of bone graphs to shock graphs in a set of view-based object recognition and pose estimation trials.

30. Automatic Non-rigid Registration of 3D Dynamic Data for Facial Expression Synthesis and Transfer

Sen Wang, Xianfeng Gu, Hong Qin

Automatic non-rigid registration of 3D time-varying data is fundamental in many vision and graphics applications such as facial expression analysis, synthesis, and recognition. Despite many research advances in recent years, it still remains to be technically challenging, especially for 3D dynamic, densely-sampled facial data with a large number of degrees of freedom (necessarily used to represent rich and subtle facial expressions). In this paper, we present a new method for automatic non-rigid registration of 3D dynamic facial data using least-squares conformal maps, and based on this registration method, we also develop a new framework of facial expression synthesis and transfer. Nowadays more and more 3D dynamic, densely-sampled data become prevalent with the advancement of novel 3D scanning techniques. To analyze and utilize such huge 3D data, an efficient non-rigid registration algorithm is needed to establish one-to-one inter-frame correspondences. Towards this goal, a non-rigid registration algorithm of 3D dynamic facial data is developed by using least-squares conformal maps with additional feature correspondences detected by employing active appearance models (AAM). The proposed method with additional, interior feature constraints guarantees that the non-rigid data will be accurately registered. The least-squares conformal maps between two 3D surfaces are globally optimized with the least angle distortion and the resulting 2D maps are stable and one-to-one. Furthermore, by using this non-rigid registration method, we develop a new system of facial expression synthesis and transfer. Finally, we perform a series of experiments to evaluate our non-rigid registration method and demonstrate its efficacy and efficiency in the applications of facial expression synthesis and transfer.

31. Accurate Multi-View Reconstruction Using Robust Binocular Stereo and Surface Meshing

Derek Bradley, Tamy Boubekeur, Wolfgang Heidrich

This paper presents a new algorithm for multi-view reconstruction that demonstrates both accuracy and efficiency. Our method is based on robust binocular stereo matching, followed by adaptive point-based filtering of the merged point clouds, and efficient, high-quality mesh generation. All aspects of our method are designed to be highly scalable with the number of views. Our technique produces the most accurate results among current algorithms for a sparse number of viewpoints according to the Middlebury datasets. Additionally, we prove to be the most efficient method among non-GPU algorithms for the same datasets. Finally, our scaled-window matching technique also excels at reconstructing deformable objects with high-curvature surfaces, which we demonstrate with a number of examples.

32. A general solution to the P4P problem for camera with unknown focal length

Martin Bujnak, Zuzana Kukelova, Tomas Pajdla

This paper presents a general solution to the determination of the pose of a perspective camera with unknown focal length from images of four 3D reference points. Our problem is a generalization of the P3P and P4P problems previously developed for fully calibrated cameras. Given four 2D-to-3D correspondences, we estimate camera position, orientation and recover the camera focal length. We formulate the problem and provide a minimal solution from four points by solving a system of algebraic equations. We compare the Hidden variable resultant and Gröbner basis techniques for solving the algebraic equations of our problem. By evaluating them on synthetic and on real-data, we show that the Gröbner basis technique provides stable results.

33. Particle Filtering for Registration of 2D and 3D Point Sets with Stochastic Dynamics

Romeil Sandhu, Samuel Dambreville, Allen Tannenbaum

In this paper, we propose a particle filtering approach for the problem of registering two point sets that differ by a rigid body transformation. Typically, registration algorithms compute the transformation parameters by maximizing a metric given an estimate of the correspondence between points across the two sets of interest. This can be viewed as a posterior estimation problem, in which the corresponding distribution can naturally be estimated using a particle filter. In this work, we treat motion as a local variation in pose parameters obtained from running a few iterations of the standard Iterative Closest Point (ICP) algorithm. Employing this idea, we introduce stochastic motion dynamics to widen the narrow band of convergence often found in local optimizer functions used to tackle the registration task. Thus, the novelty of our method is twofold: Firstly, we employ a particle filtering scheme to drive the point set registration process. Secondly, we increase the robustness of the registration performance by introducing a dynamic model of uncertainty for the transformation parameters. In contrast with other techniques, our approach requires no annealing schedule, which results in a reduction in computational complexity as well as maintains the temporal coherency of the state (no loss of information). Also, unlike most alternative approaches for point set registration, we make no geometric assumptions on the two data sets. Experimental results are provided that demonstrate the robustness of the algorithm to initialization, noise, missing structures or differing point densities in each sets, on challenging 2D and 3D registration tasks.

34. View-invariant Recognition of Body Pose from Space-Time Templates

Yuping Shen, Hassan Foroosh

We propose a new template-based approach for view-invariant recognition of body poses based on geometric constraints derived from the motion of body point triplets. In addition to spatial information, our templates encode temporal information of body pose transitions. Unlike existing methods that study a body pose as a whole, we decompose it into a number of body point triplets, and compare their motions to our templates. Using the fact that the homography induced by the motion of a triplet of body points in two identical body pose transitions reduces to the special case of a homology, we exploit the equality of two of its eigenvalues to impose constraints on the similarity of the pose transitions between two subjects, observed by different perspective cameras and from different viewpoints. Extensive experimental results show that our method can accurately identify human poses from video sequences when they are observed from totally different viewpoints with different camera parameters.

35. Conjugate Rotation: Parameterization and Estimation from an Affine Feature Correspondence

Kevin Koeser, Christian Beder, Reinhard Koch

When rotating a pinhole camera, images are related by the infinite homography KRK^{-1}, which is algebraically a conjugate rotation. Although being a very common image transformation, e.g., important for self-calibration or panoramic image mosaicing, it is not completely understood yet. We show that a conjugate rotation has 7 degrees of freedom (as opposed to 8 for a general homography) and give a minimal parameterization. To estimate the conjugate rotation, authors traditionally made use of point correspondences, which can be seen as local zero order Taylor approximations to the image transformation. Recently however, affine feature correspondences have become increasingly popular. We observe that each such affine correspondence now provides a local first order Taylor approximation, which has not been exploited in the context of geometry estimation before. Using those two novel concepts above, we finally show that it is possible to estimate a conjugate rotation from a single affine feature correspondence under the assumption of square pixels and zero skew. As a byproduct, the proposed algorithm directly yields rotation, focal length and principal point.

36. Verifying Global Minima for L_2 Minimization Problems

Richard Hartley, Yongduek Seo

We consider the least-squares (L2) triangulation problem and structure-and-motion with known rotatation, or known plane. Although optimal algorithms have been given for these algorithms under an L-infinity cost function, finding optimal least-squares (L2) solutions to these problems is difficult, since the cost functions are not convex, and in the worst case can have multiple minima. Iterative methods can usually be used to find a good solution, but this may be a local minimum. This paper provides a method for verifying whether a local-minimum solution is globally optimal, by providing a simple and rapid test involving the Hessian of the cost function. In tests of a data set involving 277,000 independent triangulation problems, it is shown that the test verifies the global optimality of an iterative solution in over 99.9% of the cases.

37. Scale Invariance without Scale Selection

Iasonas Kokkinos, Alan Yuille

In this work we construct scale invariant descriptors (SIDs) without requiring the estimation of image scale; we thereby avoid scale selection which is often unreliable.

Our starting point is a combination of Log-Polar sampling and spatially-varying smoothing that converts image scalings and rotations into translations. Scale invariance can then be guaranteed by estimating the Fourier Transform Modulus (FTM) of the formed signal as the FTM is translation invariant.

We build our descriptors using phase, orientation and amplitude features that compactly capture the local image structure. Our results show that the constructed SIDs outperform state-of-the-art descriptors on standard datasets.

A main advantage of SIDs is that they are applicable to a broader range of image structures, such as edges, for which scale selection is unreliable. We demonstrate this by combining SIDs with contour segments and show that the performance of a boundary-based model is systematically improved on an object detection task.

38. Object Categorization using Co-Occurrence, Location and Appearance

Carolina Galleguillos, Andrew Rabinovich, Serge Belongie

In this work we introduce a novel approach to object categorization that incorporates two types of context--co-occurrence and relative location--with local appearance-based features. Our approach, named CoLA (for Co-occurrence, Location and Appearance), uses a conditional random field (CRF) to maximize object label agreement according to both semantic and spatial relevance. We model relative location between objects using simple pairwise features. By vector quantizing this feature space, we learn a small set of prototypical spatial relationships directly from the data. We evaluate our results on two challenging datasets: PASCAL 2007 and MSRC. The results show that combining co-occurrence and spatial context improves accuracy in as many as half of the categories compared to using co-occurrence alone.

39. Discriminative Local Binary Patterns for Human Detection in Personal Album

Yadong Mu, Shuicheng Yan, Yi Liu, Thomas Huang, Bingfeng Zhou

In recent years, local pattern based object detection and recognition have attracted increasing interest in computer vision research community. However, to our best knowledge no previous work has focused on utilizing local patterns for the task of human detection. In this paper we develop a novel human detection system in personal albums based on LBP (local binary pattern) descriptor. Firstly we review the existing gradient based local features widely used in human detection, analyze their limitations and argue that LBP is more discriminative. Secondly, original LBP descriptor does not suit the human detecting problem well due to its high complexity and lack of semantic consistency, thus we propose two variants of LBP: Semantic-LBP and Fourier-LBP. Carefully designed experiments demonstrate the superiority of LBP over other traditional features for human detection. Especially we adopt a random ensemble algorithm for better comparison between different descriptors. All experiments are conducted on INRIA human database.

40. Taylor Expansion Based Classifier Adaptation: Application to Person Detection

Cha Zhang, Raffay Hamid, Zhengyou Zhang

Because of the large variation across different environments, a generic classifier trained on extensive data-sets may perform sub-optimally in a particular test environment. In this paper, we present a general framework for classifier adaptation, which improves an existing generic classifier in the new test environment. Viewing classifier learning as a cost minimization problem, we perform classifier adaptation by combining the cost function on the old data-sets with the cost function on the data-set collected from the new environment. The former term is further approximated with its second order Taylor expansion to reduce the amount of information that needs to be saved for adaptation. Unlike traditional approaches that are often designed for a specific application or classifier, our scheme is applicable to various types of classifiers and user labels. We demonstrate this property on two popular classifiers (logistic regression and boosting), while using two types of user labels (direct labels and similarity labels). Extensive experiments conducted for the task of person detection in conference-room environments show that significant performance improvement can be achieved with our proposed method.

41. Locally Assembled Binary (LAB) Feature for Fast and Accurate Face Detection

Shengye Yan, Shiguang Shan, Xilin Chen, Wen Gao

In this paper, we describe a novel type of feature for fast and accurate face detection. The feature is called Locally Assembled Binary (LAB) Haar feature. LAB feature is basically inspired by the success of Haar feature and Local Binary Pattern (LBP) for face detection, but it is far beyond a simple combination. In our method, Haar features are modified to keep only the ordinal relationship (named by binary Haar feature) rather than the difference between the accumulated intensities. Several neighboring binary Haar features are then assembled to capture their co-occurrence with similar idea to LBP. We show that the feature is more efficient than Haar feature and LBP both in discriminating power and computational cost. Furthermore, a novel efficient detection method called feature-centric cascade is proposed to build an efficient detector, which is developed from the feature-centric method. Experimental results on the CMU+MIT frontal face test set and CMU profile test set show that the proposed method can achieve very good results and amazing detection speed.

42. Decomposition, Discovery and Detection of Visual Categories Using Topic Models

Mario Fritz, Bernt Schiele

We present a novel method for the discovery and detection of visual object categories based on decompositions using topic models. The approach is capable of learning a compact and low dimensional representation for multiple visual categories from multiple view points without labeling of the training instances. The learnt object components range from local structures over line segments to global silhouette-like descriptions. This representation can be used to discover object categories in a totally unsupervised fashion. Furthermore we employ the representation as the basis for building a supervised multi-category detection system making efficient use of training examples and outperforming pure features-based representations. The proposed speed-ups make the system scale to large databases. Experiments on three databases show that the approach improves the state-of-the-art in unsupervised learning as well as supervised detection. In particular we improve the state-of-the-art on the challenging PASCAL'06 multi-class detection tasks for several categories.

43. Misalignment Robust Face Recognition

Huan Wang, Shuicheng Yan, Jianzhuang Liu, Xiaoou Tang, Thomas Huang

In this paper, we study the problem of subspace-based face recognition under scenarios with spatial misalignments and/or image occlusions. For a given subspace, the embedding of a new datum and the underlying spatial misalignment parameters are simultaneously inferred by solving a constrained ℓ_1 norm optimization problem, which minimizes the error between the misalignment-amended image and the image reconstructed from the given subspace along with its principal complementary subspace. A byproduct of this formulation is the capability to detect the underlying image occlusions. Extensive experiments on spatial misalignment estimation, image occlusion detection, and face recognition with spatial misalignments and image occlusions all validate the effectiveness of our proposed general formulation.

44. 3D Face Tracking and Expression Inference from a 2D Sequence using Manifold Learning

Wei-Kai Liao, Gérard Medioni

We propose a person-dependent, manifold-based approach for modeling and tracking rigid and nonrigid 3D facial deformations from a monocular video sequence. The rigid and nonrigid motions are analyzed simultaneously in 3D, by automatically fitting and tracking a set of landmarks. We do not represent all nonrigid facial deformations as a simple complex manifold, but instead decompose them on a basis of eight 1D manifolds. Each 1D manifold is learned offline from sequences of labeled expressions, such as smile, surprise, etc. Any expression is then a linear combination of values along these 8 axes, with coefficient representing the level of activation. We experimentally verify that expressions can indeed be represented this way, and that individual manifolds are indeed 1D. The manifold dimensionality estimation, manifold learning, and manifold traversal operation are all implemented in the N-D Tensor Voting framework. Using simple local operations, this framework gives an estimate of the tangent and normal spaces at every sample, and provides excellent robustness to noise and outliers. The output of our system, besides the tracked landmarks in 3D, is a labeled annotation of the expression. We demonstrate results on a number of challenging sequences.

45. Action recognition using ballistic dynamics

Shiv Vitaladevuni, Vili Kellokumpu, Larry Davis

We present a Bayesian framework for action recognition through ballistic dynamics. Psycho-kinesiological studies indicate that ballistic movements form the natural units for human movement planning. The framework leads to an efficient and robust algorithm for temporally segmenting videos into atomic movements. Individual movements are annotated with person-centric morphological labels called ballistic verbs. This is tested on a dataset of interactive movements, achieving high recognition rates. The approach is also applied on a gesture recognition task, improving a previously reported recognition rate from 84% to 92%. Consideration of ballistic dynamics enhances the performance of the popular Motion History Image feature. We also illustrate the approach's general utility on real-world videos. Experiments indicate that the method is robust to view, style and appearance variations.

46. Real-Time Face Pose Estimation from Single Range Images

Michael D. Breitenstein, Daniel Kuettel, Thibaut Weise, Luc Van Gool, Hanspeter Pfister

We present a real-time algorithm to estimate the 3D pose of a previously unseen face from a single range image. Based on a novel shape signature to identify noses in range images, we generate candidates for their positions, and then generate and evaluate many pose hypotheses in parallel using modern graphics processing units (GPUs). We developed a novel error function that compares the input range image to precomputed pose images of an average face model. The algorithm is robust to large pose variations of ±90 yaw, ±45 pitch and ±30 roll rotation, facial expression, partial occlusion, and works for multiple faces in the field of view. It correctly estimates 97.8% of the poses within yaw and pitch error of 15 at 55.8 fps. To evaluate the algorithm, we built a database of range images with large pose variations and developed a method for automatic ground truth annotation.

47. Enforcing Convexity for Improved Alignment with Constrained Local Model

Yang Wang, Simon Lucey, Jeffrey Cohn

Constrained local models (CLMs) have recently demonstrated good performance in non-rigid object alignment/tracking in comparison to leading holistic approaches (e.g., AAMs). A major problem hindering the development of CLMs further, for non-rigid object alignment/tracking, is how to jointly optimize the global warp update across all local search responses. Previous methods have either used general purpose optimizers (e.g., simplex methods) or graph based optimization techniques. Unfortunately, problems exist with both these approaches when applied to CLMs. In this paper, we propose a new approach for optimizing the global warp update in an efficient manner by enforcing convexity at each local patch response surface. Furthermore, we show that the classic Lucas-Kanade approach to gradient descent image alignment can be viewed as a special case of our proposed framework. Finally, we demonstrate that our approach receives improved performance for the task of non-rigid face alignment/tracking on the MultiPIE database and the UNBC-McMaster archive.

48. Hallucinating 3D Facial Shapes

Gang Pan, Song Han, Zhaohui Wu

This paper focuses on hallucinating a facial shape from a low-resolution 3D facial shape. Firstly, we give a constrained conformal embedding of 3D shape in R^2, which establishes an isomorphic mapping between curved facial surface and 2D planar domain. With such conformal embedding, two planar representations of 3D shapes are proposed: **Gaussian curvature image** (**GCI**) for a facial surface, and **surface displacement image** (**SDI**) for a pair of facial surfaces. The conformal planar representation reduces the data complexity from 3D irregular curved surface to 2D regular grid while preserving the necessary information for hallucination. Then, hallucinating a low resolution facial shape is formalized as inference of SDI from GCIs by modeling the relationship between GCI and SDI by RBF regression. The experiments on USF HumanID 3D face database demonstrate the effectiveness of the approach. Our method can be easily extended to hallucinate those category-specific 3D surfaces sharing with similar geometric structures.

49. Simultaneous Super-Resolution and Feature Extraction for Recognition of Low-Resolution Faces

Pablo H. Hennings-Yeomans, Simon Baker, B. V. K. Vijaya Kumar

Face recognition degrades when faces are of very low resolution since many details about the difference between one person and another can only be captured in images of sufficient resolution. In this work, we propose a new procedure for recognition of low-resolution faces, when there is a high-resolution training set available. Most previous super-resolution approaches are aimed at reconstruction, with recognition only as an after-thought. In contrast, in the proposed method, face features, as they would be extracted for a face recognition algorithm (e.g., eigenfaces, Fisherfaces, etc.), are included in a super-resolution method as prior information. This approach simultaneously provides measures of fit of the super-resolution result, from both reconstruction and recognition perspectives. This is different from the conventional paradigms of matching in a low-resolution domain, or, alternatively, applying a super-resolution algorithm to a low-resolution face and then classifying the super-resolution result. We show, for example, that recognition of faces of as low as 6×6 pixel size is considerably improved compared to matching using a super-resolution reconstruction followed by classification, and to matching with a low-resolution training set.

50. Face Illumination Normalization on Large and Small Scale Features

Xiaohua Xie, Wei-Shi Zheng, Jianhuang Lai, Pong C. Yuen

It is well known that the effect of illumination is mainly on the large-scale features (low-frequency components) of a face image. In solving the illumination problem for face recognition, most (if not all) existing methods either only use extracted small-scale features while discard large-scale features, or perform normalization on the whole image. In the latter case, small-scale features may be distorted when the large-scale features are modified. In this paper, we argue that large-scale features of face image are important and contain useful information for face recognition as well as visual quality of normalized image. Moreover, this paper suggests that illumination normalization should mainly perform on large-scale features of face image rather than the whole face image. Along this line, a novel framework for face illumination normalization is proposed. In this framework, a single face image is first decomposed into large- and small- scale feature images using logarithmic total variation (LTV) model. After that, illumination normalization is performed on large-scale feature image while small-scale feature image is smoothed. Finally, a normalized face image is generated by combination of the normalized large-scale feature image and smoothed small-scale feature image. CMU PIE and (Extended) YaleB face databases with different illumination variations are used for evaluation and the experimental results show that the proposed method outperforms existing methods.

51. Precise Detailed Detection of Faces and Facial Features

Liya Ding, Aleix Martinez

Face detection has advanced dramatically over the past three decades. Algorithms can now quite reliably detect faces in clutter in or near real time. However, much still needs to be done to provide an accurate and detailed description of external and internal features. This paper presents an approach to achieve this goal. Previous learning algorithms have had limited success on this task because the shape and texture of facial features varies widely under changing expression, pose and illumination. We address this problem with the use of subclass divisions. In this approach, we use an algorithm to automatically divide the training samples of each facial feature into a set of subclasses, each representing a distinct construction of the same facial component (e.g., close versus open eye lids). The key idea used to achieve accurate detections is to not only learn the textural information of the facial feature to be detected but that of its context (i.e., surroundings). This process permits a precise detection of key facial features. We then combine this approach with edge and color segmentation to provide an accurate and detailed detection of the shape of the major facial features (brows, eyes, nose, mouth and chin). We use this face detection algorithm to obtain precise descriptions of the facial features in video sequences of American Sign Language (ASL) sentences, where the variability in expressions can be extreme. Extensive experimental validation demonstrates our method is almost as precise as manual detection, ~2% error.

52. Recognising faces in unseen modes: a tensor based approach

Santu Rana, Wanquan Liu, Mihai M. Lazarescu, Svetha Venkatesh

This paper addresses the limitation of current multilinear techniques (multilinear PCA, multilinear ICA) when applied to face recognition for handling faces in unseen illumination and viewpoints. We propose a new recognition method, exploiting the interaction of all the subspaces resulting from multilinear decomposition (for both multilinear PCA and ICA), to produce a new basis called multilinear-eigenmodes. This basis offers the flexibility to handle face images at unseen illumination or viewpoints. Experiments on benchmarked datasets yield superior performance in terms of both accuracy and computational cost.

53. On the use of Independent Tasks for Face Recognition

Agata Lapedriza, David Masip, Jordi Vitria

We present a method for learning discriminative linear feature extraction using independent tasks. More concretely, given a target classification task, we consider a complementary classification task that is independent of the target one. For example, in face classification field, subject recognition can be a target task while facial expression classification can be a complementary task. Then, we use labels of the complementary task in order to obtain a more robust feature extraction, being the new feature space less sensitive to the complementary classification. To learn the proposed feature extraction we use the mutual information measure between the projected data and both labels from the target and the complementary tasks. In our experiments, this framework has been applied to a face recognition problem, in order to inhibit this classification task from environmental artifacts, and to mitigate the effects of the small sample size problem. Our classification experiments show an improved feature extraction process using the proposed method.

54. Cost-Sensitive Face Recognition

Yin Zhang, Zhi-Hua Zhou

Traditional face recognition systems attempt to achieve a high recognition accuracy, which implicitly assumes that the losses of all misclassifications are the same. However, in many real-world tasks this assumption is not always reasonable. For example, it will be troublesome if a face-recognition-based door-locker misclassifies a family member as a stranger such that s/he were not allowed to enter the house; but it will be a much more serious disaster if a stranger were misclassified as a family member and allowed to enter the house. In this paper, we propose a framework which formulates the problem as a multi-class cost-sensitive learning task, and propose a theoretically sound method based on Bayes decision theory to solve this problem. Experimental results demonstrate the effectiveness and efficiency of the proposed method.

55. Pair-Activity Classification by Bi-Trajectory Analysis

Yue Zhou, Shuicheng Yan, Thomas Huang

In this paper, we address the pair-activity classification problem, which explores the relationship between two active objects based on their motion information. Our contributions are three-fold. First, we design a set of features, e.g., causality ratio and feedback ratio based on the Granger Causality Test (GCT), for describing the pair-activities encoded as trajectory pairs. These features along with conventional velocity and position features are essentially of multi-modalities, and may be greatly different in scale and importance. To make full use of them, we then present a novel feature normalization procedure to learn the coefficients for weighting these features by maximizing the discriminating power measured by weighted correlation. Finally, we collected a pair-activity database of five categories, each of which consists of about 170 instances. The extensive experiments on this database validate the effectiveness of the designed features for pair-activity representation, and also demonstrate that the proposed feature normalization procedure greatly boosts the pair-activity classification accuracy.

56. Who Killed the Directed Model?

Justin Domke, Alap Karapurkar, Yiannis Aloimonos

Prior distributions are useful for robust low-level vision, and undirected models (e.g., Markov Random Fields) have become a central tool for this purpose. Though sometimes these priors can be specified by hand, this becomes difficult in large models, which has motivated learning these models from data. However, maximum likelihood learning of undirected models is extremely difficult- essentially all known methods require approximations and/or high computational cost.

Conversely, directed models are essentially trivial to learn from data, but have not received much attention for low-level vision. We compare the two formalisms of directed and undirected models, and conclude that there is no *a priori* reason to believe one better represents low-level vision quantities. We formulate two simple directed priors, for natural images and stereo disparity, to empirically test if the undirected formalism is superior. We find in both cases that a simple directed model can achieve results similar to the best learnt undirected models with significant speedups in training time, suggesting that directed models are an attractive choice for tractable learning.

57. Object image retrieval by exploiting online knowledge resources

Gang Wang, David Forsyth

We describe a method to retrieve images found on web pages with specified object class labels, using an analysis of text around the image and of image appearance. Our method determines whether an object is both described in text and appears in a image using a discriminative image model and a generative text model.

Our models are learnt by exploiting established online knowledge resources (Wikipedia pages for text; Flickr and Caltech data sets for image). These resources provide rich text and object appearance information. We describe results on two data sets. The first is Berg's collection of ten animal categories; on this data set, we outperform previous approaches [7, 33]. We have also collected five more categories. Experimental results show the effectiveness of our approach on this new data set.

58. Finding Trails

Scott Morris, Kobus Barnard

We present a statistical learning approach for finding recreational trails in aerial images. While the problem of recognizing relatively straight and well defined roadways in digital images has been well studied in the literature, the more difficult problem of extracting trails has received no attention. However, trails and rough roads are less likely to be adequately mapped, and change more rapidly over time. Automated tools for finding trails will be useful to cartographers, recreational users and governments. In addition, the methods developed here are applicable to the more general problem of finding linear structure.

Our approach combines local estimates for image pixel trail probabilities with the global constraint that such pixels must link together to form a path. For the local part, we present results using three classification techniques. To construct a global solution (a trail) from these probabilities, we propose a global cost function that includes both global probability and path length. We show that the addition of a length term significantly improves trail finding ability. However, computing the optimal trail becomes intractable as known dynamic programming methods do not apply. Thus we describe a new splitting heuristic based on Dijkstra's algorithm. We then further improve upon the results with a trail sampling scheme.

We test our approach on 500 challenging images along the 2500 mile continental divide mountain bike trail, where assumptions prevalent in the road literature are violated.

59. A Framework for Reducing Ink-Bleed in Old Documents

Yi Huang, Michael S. Brown, Dong Xu

We describe a novel application framework to reduce the effects of ink-bleed in old documents. This task is treated as a classification problem where training-data is used to compute per-pixel likelihoods for use in a dual-layer Markov Random Field (MRF) that simultaneously labels image pixels of the front and back of a document as either foreground, background, or ink-bleed, while maintaining the integrity of foreground strokes. Our approach obtains better results than previous work without the need for assumptions about ink-bleed intensities or extensive parameter tuning. Our overall framework is detailed, including front and back image alignment, training-data collection, and the MRF formulation with associated likelihoods and intra- and interlayer cost computations.

60. Nonlinear Image Representation Using Divisive Normalization

Siwei Lyu, Eero Simoncelli

In this paper, we describe a nonlinear image representation based on divisive normalization that is designed to match the statistical properties of photographic images, as well as the perceptual sensitivity of biological visual systems. We decompose an image using a multi-scale oriented representation, and use Student's t as a model of the dependencies within local clusters of coefficients. We then show that normalization of each coefficient by the square root of a linear combination of the amplitudes of the coefficients in the cluster reduces statistical dependencies. We further show that the resulting divisive normalization transform is invertible and provide an efficient iterative inversion algorithm. Finally, we probe the statistical and perceptual advantages of this image representation by examining its robustness to added noise, and using it to enhance image contrast.

61. Utilizing Semantic Word Similarity Measures for Video Retrieval

Yusuf Aytar, Mubarak Shah, Jiebo Luo

This is a high level computer vision paper, which employs concepts from Natural Language Understanding in solving the video retrieval problem. Our main contribution is the utilization of the semantic word similarity measures (Lin and PMI-IR similarities) for video retrieval. In our approach, we use trained concept detectors, and the visual co-occurrence relations between such concepts. We propose two methods for content-based retrieval of videos: (1) A method for *retrieving a **new** concept* (a concept which is not known to the system, and no annotation is available) using semantic word similarity and visual co-occurrence. (2) A method for retrieval of videos based on their relevance to a user defined text query using the semantic word similarity and visual content of videos. For evaluation purposes, we have mainly used the automatic search and the high level feature extraction test set of TRECVID'06 benchmark, and the automatic search test set of TRECVID'07. These two data sets consist of 250 hours of multilingual news video captured from American, Arabic, German and Chinese TV channels. Although our method for retrieving a new concept is an unsupervised method, it outperforms the trained concept detectors (which are supervised) on 7 out of 20 test concepts, and overall it performs very close to the trained detectors. On the other hand, our visual content based semantic retrieval method performs 81% better than the text-based retrieval method. This shows that using visual content alone we can obtain significantly good retrieval results.

62. Application and Evaluation of Spatiotemporal Enhancement of Live Aerial Video using Temporally Local Mosaics

Bryan Morse, Damon Gerhardt, Cameron Engh, Michael Goodrich, Nathan Rasmussen, Daniel Thornton

Camera-equipped mini-UAVs are popular for many applications, including search and surveillance, but video from them is commonly plagued with distracting jittery motions and disorienting rotations that make it difficult for human viewers to detect objects of interest and infer spatial relationships. For time-critical search situations there are also inherent tradeoffs between detection and search speed. These problems make the use of dynamic mosaics to expand the spatiotemporal properties of the video appealing. However, for many applications it may not be necessary to maintain full mosaics of all of the video but to mosaic and retain only a number of recent (temporally local) frames, still providing a larger field of view and effectively longer temporal view as well as natural stabilization and consistent orientation. This paper presents and evaluates a real-time system for displaying live video to human observers in search situations by using temporally local mosaics while avoiding masking effects from dropped or noisy frames. Its primary contribution is an empirical study of the effectiveness of using such methods for enhancing human detection of objects of interest, which shows that temporally local mosaics increase task performance and are easier for humans to use than non-mosaiced methods, including stabilized video.

63. Performance Evaluation of State-of-the-Art Discrete Symmetry Detection Algorithms

Minwoo Park, Seungkyu Lee, Po-Chun Chen, Somesh Kashyap, Asad Butt, Yanxi Liu

Symmetry is one of the important cues for human and machine perception of the world. For over three decades, automatic symmetry detection from images/patterns has been a standing topic in computer vision. We present a timely, systematic, and quantitative performance evaluation of three state of the art discrete symmetry detection algorithms. This evaluation scheme includes a set of carefully chosen synthetic and real images presenting justified, unambiguous single or multiple dominant symmetries, and a pair of well-defined success rates for validation. We make our 176 test images with associated hand-labeled ground truth publicly available with this paper. In addition, we explore the potential contribution of symmetry detection for object recognition by testing the symmetry detection algorithm on three publicly available object recognition image sets (PASCAL VOC'07, MSRC and Caltech-256). Our results indicate that even after several decades of effort, symmetry detection in real-world images remains a challenging, unsolved problem in computer vision. Meanwhile, we illustrate its future potential in object recognition.

64. Enhancing Photographs with Near Infrared Images

Xiaopeng Zhang, Terence Sim, Xiaoping Miao

Near Infra-Red (NIR) images of natural scenes usually have better contrast and contain rich texture details that may not be perceived in visible light photographs (VIS). In this paper, we propose a novel method to enhance a photograph by using the contrast and texture information of its corresponding NIR image. More precisely, we first decompose the NIR/VIS pair into average and detail wavelet subbands. We then transfer the contrast in the average subband and transfer texture in the detail subbands. We built a special camera mount that optically aligns two consumer-grade digital cameras, one of which was modified to capture NIR. Our results exhibit higher visual quality than tone-mapped HDR images, showing that NIR imaging is useful for computational photography.

65. Vital Sign Estimation from Passive Thermal Video

Ming Yang, Qiong Liu, Thea Turner, Ying Wu

Conventional wired detection of vital signs limits the use of these important physiological parameters by many applications, such as airport health screening, elder care, and workplace preventive care. In this paper, we explore contact-free heart rate and respiratory rate detection through measuring infrared light modulation emitted near superficial blood vessels or a nasal area respectively. To deal with complications caused by subjects' movements, facial expressions, and partial occlusions of the skin, we propose a novel algorithm based on contour segmentation and tracking, clustering of informative pixels, and dominant frequency component estimation. The proposed method achieves robust subject regions-of-interest alignment and motion compensation in infrared video with low SNR. It relaxes some strong assumptions used in previous work and substantially improves on previously reported performance. Preliminary experiments on heart rate estimation for 20 subjects and respiratory rate estimation for 8 subjects exhibit promising results.

66. What Are the High-Level Concepts with Small Semantic Gaps?

Yijuan Lu, Lei Zhang, Qi Tian, Weiying Ma

Concept-based multimedia search has become more and more popular in Multimedia Information Retrieval (MIR). However, which semantic concepts should be used for data collection and model construction is still an open question. Currently, there is very little research found on automatically choosing multimedia concepts with small semantic gaps. In this paper, we propose a novel framework to develop a lexicon of high-level concepts with small semantic gaps (LCSS) from a large-scale web image dataset. By defining a confidence map and content-context similarity matrix, images with small semantic gaps are selected and clustered. The final concept lexicon is mined from the surrounding descriptions (titles, categories and comments) of these images. This lexicon offers a set of high-level concepts with small semantic gaps, which is very helpful for people to focus for data collection, annotation and modeling. It also shows a promising application potential for image annotation refinement and rejection. The experimental results demonstrate the validity of the developed concepts lexicon.

67. Extrinsic and Depth Calibration of ToF-cameras

Stefan Fuchs, Gerd Hirzinger

Recently, ToF-cameras have attracted attention because of their ability to generate a full 2½ *D* depth image at video frame rates. Thus, ToF-cameras are suitable for real-time 3D tasks such as tracking, visual servoing or object pose estimation. The usability of such systems mainly depends on an accurate camera calibration. In this work a calibration process for ToF-cameras with respect to the intrinsic parameters, the depth measurement distortion and the pose of the camera relative to a robot's endeffector is described. The calibration process is not only based on the monochromatic images of the camera but also uses its depth values that are generated from a chequer-board pattern. The robustness and precision of the presented method is assessed applying it to randomly selected shots and comparing the calibrated measurements to a ground truth obtained from a laser scanner.

68. Localization Accuracy of Region Detectors

Andreas Haja, Steffen Abraham, Bernd Jaehne

In this paper, a comparison of five state of the art region detectors is presented with regard to localization accuracy in position and region shape. Based on carefully estimated ground truth homographies, correspondences between frames are assigned using geometrical region overlap. Significant differences between detectors exist, depending on the type of images. Also, it is shown that localization accuracy linearly depends on region scale for some detectors, which may thus be used as a pre-selection criterion for the removal of error-prone regions. The presented results serve as a supplement to existing comparative studies, and can be used to facilitate the selection of an appropriate detector for a specific target application. When descriptor distance is used as assignment criterion instead of region overlap, a different set of correspondences results with lower accuracy. Set differences (and thus localization accuracy) are directly related to the density of regions in a local neighborhood. Based on the latter, a novel measure for the identification of error-prone regions - shape uniqueness - is introduced. In contrast to existing methods that are based on the descriptor distance of region correspondences, the new measure is pre-computed on each image individually. Thus, the complexity of the subsequent matching task can be significantly reduced.

69. Robust Dual Motion Deblurring

Jia Chen, Lu Yuan, Chi-Kueng Tang, Long Quan

This paper presents a robust algorithm to deblur two consecutively captured blurred photos from camera shaking. Previous dual motion deblurring algorithms succeeded in small and simple motion blur and are very sensitive to noise. We develop a robust feedback algorithm to perform iteratively kernel estimation and image deblurring. In kernel estimation, the stability and capability of the algorithm is greatly improved by incorporating a robust cost function and a set of kernel priors. The robust cost function serves to reject outliers and noise, while kernel priors, including sparseness and continuity, remove ambiguity and maintain kernel shape. In deblurring, we propose a novel and robust approach which takes two blurred images as input to infer the clear image. The deblurred image is then used as feedback to refine kernel estimation. Our method can successfully estimate large and complex motion blurs which cannot be handled by previous dual or single image motion deblurring algorithms. The results are shown to be significantly better than those of previous approaches.

70. Rotation Symmetry Group Detection Via Frequency Analysis of Frieze-Expansions

Seungkyu Lee, Robert Collins, Yanxi Liu

We present a novel and effective algorithm for rotation symmetry group detection from real-world images. We propose a frieze-expansion method that transforms rotation symmetry group detection into a simple translation symmetry detection problem. We define and construct a dense symmetry strength map from a given image, and search for potential rotational symmetry centers automatically. Frequency analysis, using Discrete Fourier Transform (DFT), is applied to the frieze-expansion patterns to uncover the types and the cardinality of multiple rotation symmetry groups in an image, concentric or otherwise. Furthermore, our detection algorithm can discriminate discrete versus continuous and cyclic versus dihedral symmetry groups, and identify the corresponding supporting regions in the image. Experimental results on over 80 synthetic and natural images demonstrate superior performance of our rotation detection algorithm in accuracy and in speed over the state of the art rotation detection algorithms.

71. Accurate and Robust Registration for In-hand Modeling

Thibaut Weise, Bastian Leibe, Luc Van Gool

We present fast 3D surface registration methods for in-hand modeling. This allows users to scan complete objects swiftly by simply turning them around in front of the scanner. The paper makes two main contributions. First, we propose an efficient method for detecting registration failures, which is a vital property of any automatic modeling system. Our method is based on two different consistency tests, one based on geometry and one based on texture. Second, we extend ICP by three additional fast registration methods for both coarse and fine alignment based on both texture and geometry. Each of those methods brings in additional information that can compensate for ambiguities in the other cues. Together, they allow for the robust reconstruction of a large variety of objects with different geometric and photometric properties. Finally, we show how both failure detection and fast registration can be combined in a practical and robust in-hand modeling system that operates at interactive frame rates.

72. Sensing Increased Image Resolution Using Aperture Masks

Ankit Mohan, Xiang Huang, Ramesh Raskar, Jack Tumblin

We present a technique to construct increased-resolution images from multiple photos taken without moving the camera or the sensor. Like other super-resolution techniques, we capture and merge multiple images, but instead of moving the camera sensor by sub-pixel distances for each image, we change masks in the lens aperture and slightly defocus the lens. The resulting capture system is simpler, and tolerates modest mask registration errors well. We present a theoretical analysis of the camera and image merging method, show both simulated results and actual results from a crudely modified consumer camera, and compare its results to robust `blind' methods that rely on uncontrolled camera displacements.

73. PSF Estimation using Sharp Edge Prediction

Neel Joshi, Richard Szeliski, David Kriegman

Image blur is caused by a number of factors such as motion, defocus, capturing light over the non-zero area of the aperture and pixel, the presence of anti-aliasing filters on a camera sensor, and limited sensor resolution. We present an algorithm that estimates non-parametric, spatially-varying blur functions (i.e., point-spread functions or PSFs) at sub-pixel resolution from a single image. Our method handles blur due to defocus, slight camera motion, and inherent aspects of the imaging system. Our algorithm can be used to measure blur due to limited sensor resolution by estimating a sub-pixel, super-resolved PSF even for in-focus images. It operates by predicting a "sharp" version of a blurry input image and uses the two images to solve for a PSF. We handle the cases where the scene content is unknown and also where a known printed calibration target is placed in the scene. Our method is completely automatic, fast, and produces accurate results.

74. Non Refractive Modulators for Encoding and Capturing Scene Appearance and Depth

Ashok Veeraraghavan, Amit Agrawal, Ramesh Raskar, Ankit Mohan, Jack Tumblin

We analyze the modulation of a light field via non-refracting attenuators. In the most general case, any desired modulation can be achieved with attenuators having four degrees of freedom in ray-space. We motivate the discussion with a universal 4D ray modulator (ray-filter) which can attenuate the intensity of each ray independently. We describe operation of such a fantasy ray-filter in the context of altering the 4D light field incident on a 2D camera sensor.

Ray-filters are difficult to realize in practice but we can achieve reversible encoding for light field capture using patterned attenuating mask. Two mask-based designs are analyzed in this framework. The first design closely mimics the angle-dependent ray-sorting possible with the ray filter. The second design exploits frequency-domain modulation to achieve a more efficient encoding. We extend these designs for optimal sampling of light field by matching the modulation function to the specific shape of the band-limit frequency transform of light field. We also show how a hand-held version of an attenuator based light field camera can be built using a medium-format digital camera and an inexpensive mask.

75. Modulated Phase-shifting for 3D Scanning

Tongbo Chen, Hans-Peter Seidel, Hendrik P. A. Lensch

We present a new 3D scanning method using modulated phase-shifting. Optical scanning of complex objects or scenes with significant global light transport, such as subsurface scattering, interreflections, volumetric scattering, etc. is a difficult task since the direct surface reflection will be mixed with the global illumination. The direct and global components can be efficiently separated using high frequency illumination which to some extend is done in traditional phase-shifting for 3D scanning. In this paper we introduce the concept of modulation based separation where a high frequency signal is multiplied on top of other signal. The modulated signal inherits the good separation properties of the high frequency signal and allows for removing artifacts due to global illumination. This technique can be used to clean up arbitrary projected signals, e.g., photographs as well as the sinusoid patterns used for phase-shifting. For the modulated phase-shifting, we propose a two-pass separation method exploiting high frequency patterns in two-dimensions that can filter out the global components much more completely than traditional one-pass separation methods. We demonstrate the effectiveness of our approach on a couple of scenes with significant subsurface scattering and interreflections.

3:45pm – 5:30pm Oral Session O3P-1: Video Analysis and Image and Video Retrieval (Cook)

1. Using Circular Statistics for Trajectory Shape Analysis

Andrea Prati, Simone Calderara, Rita Cucchiara

The analysis of patterns of movement is a crucial task for several surveillance applications, for instance to classify normal or abnormal people trajectories on the basis of their occurrence. This paper proposes to model the shape of a single trajectory as a sequence of angles described using a Mixture of Von Mises (MoVM) distribution. A complete EM (Expectation Maximization) algorithm is derived for MoVM parameters estimation and an on-line version proposed to meet real time requirement. Maximum-A-Posteriori is used to encode the trajectory as a sequence of symbols corresponding to the MoVM components. Iterative k-medoids clustering groups trajectories in a variable number of similarity classes. The similarity is computed aligning (with dynamic programming) two sequences and considering as symbol-to-symbol distance the Bhattacharyya distance between von Mises distributions. Extensive experiments have been performed on both synthetic and real data.

2. Shape L'Âne Rouge: Sliding Wavelets for Indexing and Retrieval

Adrian M. Peter, Anand Rangarajan, Jeffrey Ho

Shape representation and retrieval of stored shape models are becoming increasingly more prominent in fields such as medical imaging, molecular biology and remote sensing. We present a novel framework that directly addresses the necessity for a rich and compressible shape representation, while simultaneously providing an accurate method to index stored shapes. The core idea is to represent point-set shapes as the square root of probability densities expanded in a wavelet basis. We then use this representation to develop a natural similarity metric that respects the geometry of these probability distributions, i.e., under the wavelet expansion, densities are points on a unit hypersphere and the distance between densities is given by the separating arc length. The process uses a linear assignment solver for non-rigid alignment between densities prior to matching; this has the connotation of "sliding" wavelet coefficients akin to the sliding block puzzle L'Âne Rouge. We illustrate the utility of this framework by matching shapes from the MPEG-7 data set and provide comparisons to other similarity measures, such as Euclidean distance shape distributions.

3. One Step Beyond Histograms: Image Representation using Markov Stationary Features

Jianguo Li, Weixin Wu, Tao Wang, Yimin Zhang

This paper proposes a general framework called Markov stationary features (MSF) to extend histogram based features. The MSF characterizes the spatial co-occurrence of histogram patterns by Markov chain models, and finally yields a compact feature representation through Markov stationary analysis. Therefore, the MSF goes one step beyond histograms since it now involves spatial structure information of both within histogram bins and between histogram bins. Moreover, it still keeps simplicity, compactness, efficiency, and robustness. We demonstrate how the MSF is used to extend histogram based features like color histogram, edge histogram, local binary pattern histogram and histogram of oriented gradients. We evaluate the MSF extended histogram features on the task of TRECVID video concept detection. Results show that the proposed MSF extensions can achieve significant performance improvement over corresponding histogram features.

4. Graph commute times for image representation

Regis Behmo, Nikos Paragios, Veronique Prinet

We introduce a new image representation that encompasses both the general layout of groups of quantized local invariant descriptors as well as their relative frequency. A graph of interest points clusters is constructed and we use the matrix of commute times between the different nodes of the graph to obtain a description of their relative arrangement that is robust to large intra class variation.

The obtained high dimensional representation is then embedded in a space of lower dimension by exploiting the spectral properties of the graph made of the different images. Classification tasks can be performed in this embedding space. We expose classification and labelling results obtained on three different datasets, including the challenging PASCAL VOC2007 dataset. The performances of our approach compare favorably with the standard bag of features, which is a particular case of our representation.

5. Fast Image Search for Learned Metrics

Prateek Jain, Brian Kulis, Kristen Grauman

We introduce a method that enables scalable image search for learned metrics. Given pairwise similarity and dissimilarity constraints between some images, we learn a Mahalanobis distance function that captures the images' underlying relationships well. To allow sub-linear time similarity search under the learned metric, we show how to encode the learned metric parameterization into randomized locality-sensitive hash functions. We further formulate an indirect solution that enables metric learning and hashing for vector spaces whose high dimensionality make it infeasible to learn an explicit weighting over the feature dimensions. We demonstrate the approach applied to a variety of image datasets. Our learned metrics improve accuracy relative to commonly-used metric baselines, while our hashing construction enables efficient indexing with learned distances and very large databases.

3:45pm – 5:30pm Oral Session O3P-2: Selected Topics (La Perouse)

1. Summarizing Visual Data Using Bidirectional Similarity

Denis Simakov, Yaron Caspi, Eli Shechtman, Michal Irani

We propose a principled approach to summarization of visual data (images or video) based on optimization of a well-defined similarity measure. The problem we consider is *re-targeting* (or summarization) of image/video data into smaller sizes. A good "visual summary" should satisfy two properties: (1) it should contain as much as possible visual information from the input data; (2) it should introduce as few as possible new visual artifacts that were not in the input data (i.e., preserve visual coherence). We propose a bi-directional similarity measure which quantitatively captures these two requirements: Two signals *S* and *T* are considered visually similar if all patches of *S* (at multiple scales) are contained in *T*, and vice versa.

The problem of summarization/re-targeting is posed as an optimization problem of this bi-directional similarity measure. We show summarization results for image and video data. We further show that the same approach can be used to address a variety of other problems, including automatic cropping, completion and synthesis of visual data, image collage, object removal, photo reshuffling and more.

2. Constant Time O(1) Bilateral Filtering

Fatih Porikli

This paper presents three novel methods that enable bilateral filtering in constant time O(1) without sampling. Constant time means that the computation time of the filtering remains same even if the filter size becomes very large. Our first method takes advantage of the integral histograms to avoid the redundant operations for bilateral filters with box spatial and arbitrary range kernels. For bilateral filters constructed by polynomial range and arbitrary spatial filters, our second method provides a direct formulation by using linear filters of image powers without any approximation. Lastly, we show that Gaussian range and arbitrary spatial bilateral filters can be expressed by Taylor series as linear filter decompositions without any noticeable degradation of filter response. All these methods drastically decrease the computation time by cutting it down constant times (e.g., to 0.06 seconds per 1MB image) while achieving very high PSNR's over 45dB. In addition to the computational advantages, our methods are straightforward to implement.

3. Flat Refractive Geometry

Tali Treibitz, Yoav Schechner, Hanumant Singh

While the study of geometry has mainly concentrated on single-viewpoint (SVP) cameras, there is growing attention to more general non-SVP systems. Here we study an important class of systems that inherently have a non-SVP: a perspective camera imaging through an interface into a medium. Such systems are ubiquitous: they are common when looking into water-based environments. The paper analyzes the common flat-interface class of systems. It characterizes the locus of the viewpoints (caustic) of this class, and proves that the SVP model is invalid in it. This may explain geometrical errors encountered in prior studies. Our physics-based model is parameterized by the distance of the lens from the medium interface, beside the focal length. The physical parameters are calibrated by a simple approach that can be based on a single-frame. This directly determines the system geometry. The calibration is then used to compensate for modeled system distortion. Based on this model, geometrical measurements of objects are significantly more accurate, than if based on an SVP model. This is demonstrated in real-world experiments.

4. Human-Assisted Motion Annotation

Ce Liu, William T. Freeman, Edward Adelson, Yair Weiss

Obtaining ground-truth motion for arbitrary, real-world video sequences is a challenging but important task for both algorithm evaluation and model design. Existing ground-truth databases are either synthetic, such as the Yosemite sequence, or limited to indoor, experimental setups, such as the database developed in Baker et al. We propose a human-in-loop methodology to create a ground-truth motion database for the videos taken with ordinary cameras in both indoor and outdoor scenes, using the fact that human beings are experts at segmenting objects and inspecting the match between two frames. We designed an interactive computer vision system to allow a user to efficiently annotate motion. Our methodology is cross-validated by showing that human annotated motion is repeatable, consistent across annotators, and close to the ground truth obtained by Baker et al. Using our system, we collected and annotated 10 indoor and outdoor real-world videos to form a ground-truth motion database. The source code, annotation tool and database is online for public evaluation and benchmarking.

5. Globally Optimal Bilinear Programming for Computer Vision Applications

Manmohan Chandraker, David Kriegman

We present a practical algorithm that provably achieves the global optimum for a class of bilinear programs commonly arising in computer vision applications. Our approach relies on constructing tight convex relaxations of the objective function and minimizing it in a branch and bound framework. A key contribution of the paper is a novel, provably convergent branching strategy that allows us to solve large-scale problems by restricting the branching dimensions to just one set of variables constituting the bilinearity.

Experiments with synthetic and real data validate our claims of optimality, speed and convergence. We contrast the optimality of our solutions with those obtained by a traditional singular value decomposition (SVD) approach. Among several potential applications, we discuss two: exemplar-based face reconstruction and non-rigid structure from motion. In both cases, we compute the best bilinear fit that represents a shape, observed in a single image from an arbitrary viewpoint, as a combination of the elements of a basis.