A Scene Recognition Algorithm Based on Covariance Descriptor

Yinghui Ge
Faculty of Information Science and Technology
Ningbo University
Ningbo, China
gyhzd@tom.com

Jianjun Yu
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai, China
onlysmooth@gmail.com

***Abstract*—Covariance descriptor which is a good method for object detection represents a region of interest by using the covariance of d-features. Different scene images are regarded as textures with different labels in this paper, and then a novel scene recognition algorithm based on covariance descriptor is proposed. Firstly, the covariance feature of scene image is retrieved. Secondly, Support Vector Machine is employed for training the combination classifiers which recognize the scene images. The experimental results demonstrate the feasibility and validity of the proposed algorithm.**

***Index Terms*—covariance descriptor, SVM, scene recognition**

I. Introduction

Scene recognition is more and more hot research issue in computer vision. To classify images into semantic types of scenes is a classical pattern recognition application in image understanding. For example, a variety of digital products, such as Digital Camera and Digital Video, become daily necessities. If these machines can classify the scenes automatically during shooting, a friendly usage and higher performance can be obtained.

Scene recognition focus on mapping a set of low-level image feature to semantically meaningful categories using classifier. So the first problem in scene recognition is to extract a good low-level image feature descriptor, such as color and texture feature. Image histogram[7] which describes the color distribution of image is a classic method for image feature. Haralick *et al.* designs a texture descriptor for image classificatio and has been adopted in the area of image retrieval and image slicing&stretching[10]. Tuzel *et al.*[6] present a texture classification method which consider texture feature as several covariance feature matrix extracted several random square regions from texture image with random sizes.

More researches on scene recognition are implemented in these years. Serrano *et al.*[2] propose an approach using simplified low-level feature set to predict multiple semantic scene attributes which are integrated by a Bayesian network. They first find and label well segmented image regions, such as grass, sky, water and then to extract the low-level color and wavelet texture feature, to learn the spatial contextual model among regions by support vector machine. Oliva and Torralba[3] provide a computational model which is based on a set of perceptual dimensions, so-called Spatial Envelope to represent the dominant spatial structure of a scene which bypasses the segmentation and the processing of individual objects or regions. The image segmentation is necessary of these methods mentioned above. Lu *et al.*[1] claimed that there is no need to segment image regions explicitly for scene classification and present a two level approach for scene recognition. They first learn mixture models based on color and texture information and produce probability density response maps(PDRM), and then train a bagged linear discriminant analysis(LDA) classifier for several scenes over those PDRMs.

In this paper, we extend the covariance descriptor to represent the global feature of image. There are two steps in our scene recognition technique. Firstly, the covariance description matrix of scene image is computed, and then a corresponding feature vector is constructed from the matrix. Secondly, a multiple class support vector machine is employed for learning the combination classifiers which recognize the scene images, and voting mechanism[17] is used in decision procedure. The corresponding experiments validate the proposed algorithm.

II. Image Representation Based on Covariance Descriptor

A. Covariance descriptor

Tuzel *et al.*[6] describes region covariance for characterizing regions. The covariance descriptor capture the different types of features' statistical properties as well as their correlation within the same representation and its dimensionality is small. We represent an image as the covariance matrix of features as illustrated in Fig. 1.

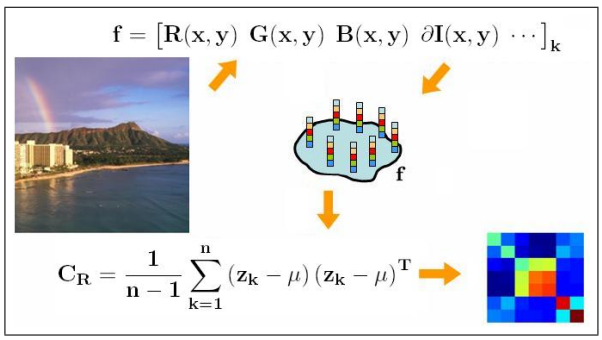


Fig. 1. Visulization of covariance feature of image.

Let I be an image with one dimensional intensity or three dimensional color. Let F be the $W \times H \times d$ dimensional features image extracted from I, shown in Equ. 1. W and H are the width and height of image respectively. d is the dimension of selected feature vector.

$$\mathbf{F}(\mathbf{x}, \mathbf{y}) = \mathbf{\Phi}(\mathbf{I}; \mathbf{x}, \mathbf{y}) \tag{1}$$

where, $\mathbf{\Phi}$ can be any mapping such as intensity, color, gradients etc. For a given region $R \subset I$ in image, let $\{z_k\}_{k=1,\ldots,n}$ be the d-dimensional feature vectors inside R. Then, a $d \times d$ covariance matrix of feature vectors is employed to represent the region R. Matrix of feature vectors C_R is

$$\mathbf{C_R} = \frac{\mathbf{1}}{\mathbf{n-1}} \sum_{\mathbf{k=1}}^{\mathbf{n}} (\mathbf{z_k} - \mu)(\mathbf{z_k} - \mu)^{\mathbf{T}} \tag{2}$$

where, μ is the mean of feature vectors. The covariance $\mathbf{C_R}$ has not any information regarding the ordering and the number of points, which is a favorable feature in the application of scene recognition.

The covariance matrix present a natural way of fusing multiple features which might be correlated. The diagonal entries of the covariance matrix represent the variance of each feature and the non-diagonal entries represent the correlations. The noise corrupting individual samples are largely filtered out with an average filter during covariance computation.

B. Image covariance feature representation

In this paper, we extend the region R to the whole image I, and an image is represented by the features covariance matrix. The pixel color(RGB) values and the norm of the first and second order derivatives of the intensities with respect to x and y are regarded as the selected features. So, each pixel of the image $\mathbf{F}$ is converted to a seven-dimensional feature vector

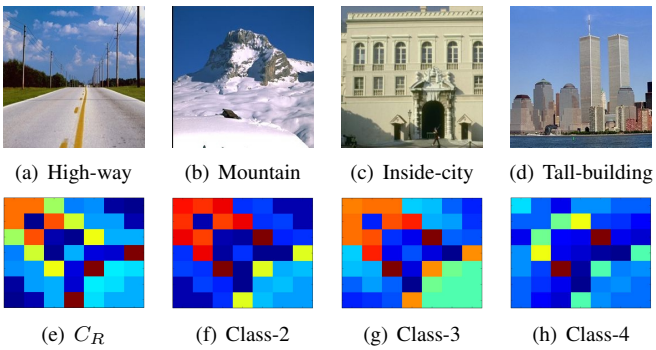
$$\mathbf{F} = \left[\mathbf{R}, \mathbf{G}, \mathbf{B}, \left|\frac{\partial \mathbf{I}}{\partial \mathbf{x}}\right|, \left|\frac{\partial \mathbf{I}}{\partial \mathbf{y}}\right|, \left|\frac{\partial^{\mathbf{2}} \mathbf{I}}{\partial^{\mathbf{2}} \mathbf{x}}\right|, \left|\frac{\partial^{\mathbf{2}} \mathbf{I}}{\partial^{\mathbf{2}} \mathbf{y}}\right|\right] \tag{3}$$

where R, G, B are the RGB color values, and I is the intensity. The image derivatives are calculated through Sobel filter. Thus, an image corresponds to a 7×7 covariance matrix which describes the variance of feature and the correlations between features. Fig. 2 shows the visualization of image features covariance matrix. The covariance matrix has only $(d^2 + d)/2$ different values due to symmetry.

III. Scene recognition based on support vector machines

The combination of kernel based feature space and hyperplane classifier builds the so-called *Support Vector Machine*(SVM)[11]. Basically they can be used to solve two-class learning problems. Some improvements compared to linear classifiers are robustness in cause of generalization and a relatively low computational cost even for large learning problems[14]. Also, there are some multiple class SVM classifiers.

All SVM algorithm present some common ideas and structures. The principle tasks are the selection of a suitable kernel



(a) High-way (b) Mountain (c) Inside-city (d) Tall-building

(e) C_R (f) Class-2 (g) Class-3 (h) Class-4

Fig. 2. Image covariance feature visualization.

to represent the data in the feature space and the computation of separating hyperplane in order to minimize the empirical risk of miss-classification.

A. Binary classifier

The simplest binary classifier of SVM is the so-called *maximal margin classifier*. It is not usable for most real-world learning applications, because it works only for data which are linearly separable in the feature space. Nonetheless it is the easiest algorithm to understand, and it forms the main building block for the more complex SVM, such as *soft margin classifier*. It exhibits the key features that characterize this kind of learning machine.

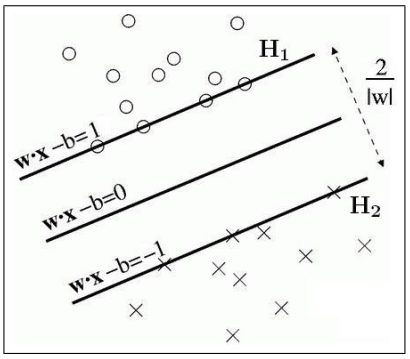


Fig. 3. Linear separating hyperplane for the separable data.

The idea of the maximal margin classifier is to separate two classes in feature space by computing the maximal margin hyperplane. This is done by minimizing a quadratic function under linear inequality constraints. Let $X = \{(x_i, c_i)\}$, $i = 1 \cdots n$, $c_i \in \{-1, +1\}$, $x_i \in R^d$. Suppose that we have some hyperplane which separates the oppositive from the negative examples, shown in Fig. 3. And suppose that all the training data satisfy the following constraints, shown in Equ. 4.

$$\begin{cases} x_i \cdot w + b \geq +1 & \text{for } c_i = +1 \\ x_i \cdot w + b \leq -1 & \text{for } c_i = -1 \end{cases} \tag{4}$$

The hyperplane can be retrieved by solving the optimization problem

$$cost\_function = \arg\min_{w,b} \|w\| \tag{5}$$

subject to

$$c_i\left(x_i \cdot w + b\right) - 1 \geq 0 \qquad \forall i \tag{6}$$

Data samples lie on the hyperplane $H_i \Rightarrow x_i \cdot w + b = 1$ and $H_2 \Rightarrow x_i \cdot w + b = -1$ are called *Support Vectors*.

Soft margin classifier which is based on the same idea as maximal margin classifier is used for noisy linearly separable training data samples and has much higher robustness. The slack variable ξ[15] is introduced to soft margin classifier, so the optimization problem is represented by

$$cost\_function = \arg\min_{w,b,\xi} \|w\| + C\sum_{i=1}^{n}\xi_i^2 \tag{7}$$

subject to

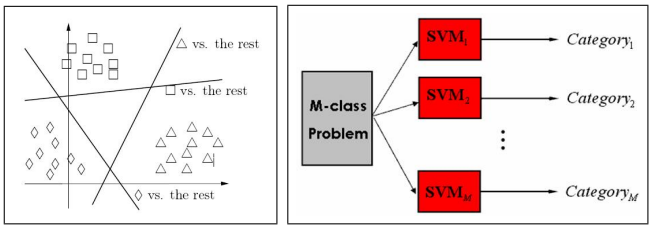
$$c_i\left(x_i \cdot w + b\right) - 1 \geq \xi_i \qquad \forall i \tag{8}$$

Where, C is a constant variable. The slack variables ξ_i are individual for each sample x_i.

B. Multiple class classifier in scene recognition

It is necessary to extend binary SVM classifier to multiple-class problems. Allwein *et al.*[16] gives a nice overview about ideas of multiple class reduction to binary problems. Sun *et al.*[17] describe two kinds weight voting schemes: classifier-based and sample-based voting strategy, and then apply them to multiple classifiers generated by Adaboost. A One-vs-Rest multiple class SVM classifier during the learning process and a voting mechanism during the decision process are used in this paper.

A natural way to solve a n-class learning problem is to into N binary learning problems. The basic idea is to formulate the problem differently: instead of learning "class 1 against class 2 against class 3 ...", the problem can be written "class 1 against the rest, class 2 against the rest, ...". Finally n binary learning problems "class i against the rest" are equivalent. The reduction to binary problems can be interpreted geometrically as searching N separating hyperplane. Fig. 4(a) shows a simple



(a) Three classes problems (b) Multiple classes framework

Fig. 4. Multiple class problems reduction.

representation of this idea and Fig. 4(b) shows a framework of multiple class problems using One-vs-Rest method.

Classifying an object in a multiple class problem is not as easy as in a true binary problem. The decision function has to be much more complex, a simple distance measure is not possible. A class-based voting mechanism[17] is employed to decide the final label of category.

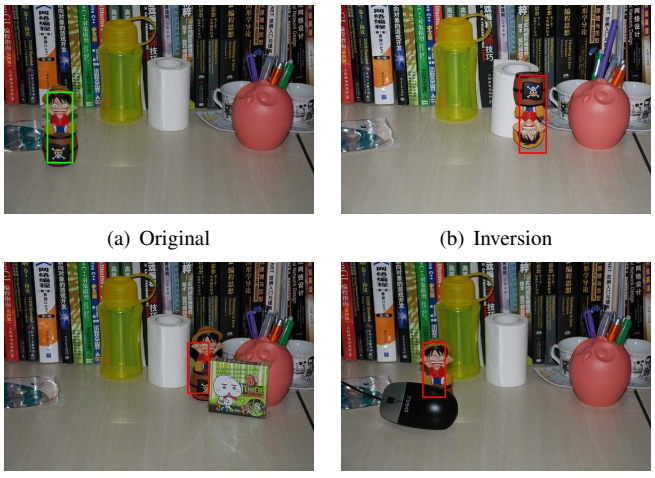
As mentioned above, scene images are mapped to a feature space by covariance descriptor. Taking into account the symmetry of covariance matrix, We consider $(d^2+d)/2 = 28$, here $d = 7$, elements in up-triangle covariance matrix as the feature vector for SVM classifiers. By using One-vs-Rest method, we reduce N, here $N = 4$, classes scene recognition issue to 4 binary classes problem. In decision process, the binary classifiers vote the scene image's label, and we hold the most votes label as the scene image's final label.

IV. Implementation

We implement the proposed algorithm using C++, OpenCV and libsvm library, and evaluate empirical performance of covariance descriptor in object detection and scene recognition.

A. Object detection

Given an object image, the object detection algorithm can locate the object in an arbitrary image. The robustness of detection mainly rely on the powerful of object appearance descriptor. The experiment results shown in Fig. 5 validate the robustness of covariance descriptor. Covariance features can match the specialized target accurately with challenging since there are orientation changes and some of the target are occluded. We conclude that the covariance descriptor has high performance in object appearance representation.



(a) Original (b) Inversion

(c) Occlusion (d) Occlusion

Fig. 5. Object detection results.

B. Scenes recognition

The complete diagram of our proposed scene recognition system is shown in Fig. 6. There are two procedures in the system, training and testing process which are described by red arrows and blue arrows separately. A combination classifiers using SVM are learned in training process including covariance descriptor, feature vectorization and multiple classes SVM. In testing process, the class labels are extracted using the combination SVM classifiers, and a voting mechanism is applied to the labels.

The scenes images set used in our experiment consists of 4 categories(High-way, Mountain, Inside-city and Tall-building)

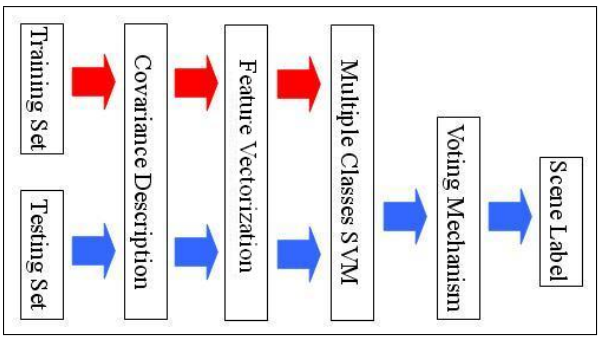
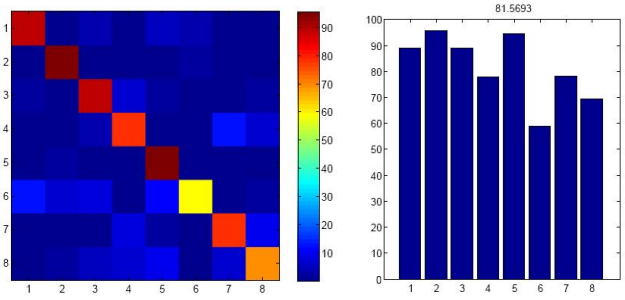


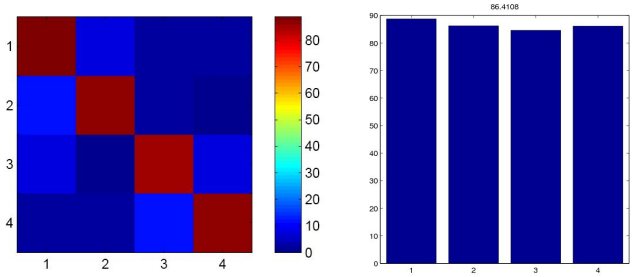
Fig. 6. Framework of Scene Recognition System.



(a) Decision matrix (b) Recognition rate

Fig. 8. Result of our method(**8** Categories).

scenes each has 400∼500 images. We randomly select 70% of them as the training data and use other images as the testing data. Fig. 7(a) shows the scene recognition decision matrix of our method. The diagonal entries of the matrix represent the rate of each SVM classifier and the non-diagonal entries represent the rate correlations between SVM classifiers. Fig. 7(b) shows the scene recognition rate of our method. The rate varies from 85%∼89%, and the mean rate is 86.4%.



(a) Decision matrix (b) Recognition rate

Fig. 7. Result of our method(**4** Categories).

We also validate the proposed method on a 8 categories scenes image set(Coast, Forest, High-way, Inside-city, Mountain, Open-country, Street and Tall-building). 80% of each scene images are randomly selected as training data and the rest are regarded as testing samples. Fig. 8 shows the experimental results. The recognition rate varies from 59.0%∼95.7%, and the mean rate is 81.6%. Compared to the result shown above, there is a large variance among the recognition rate and a depressed mean rate. With a view to the increased categories, the depressed performance is reasonable.

Oliva and Torralba[3] obtained 83% recognition rate on the same 8 categories scenes image set by modeling the shape of scene. By a simple and less computation, we attain almost recognition rate as Oliva's. Our experimental results demonstrate the validity of the proposed algorithm.

V. Conclusion

A novel scene recognition algorithm based on covariance descriptor is proposed in this paper. As a good method for object detection, covariance descriptor is to represent a region of interest by using the covariance of d-features[6]. We regard different scene images as textures with different labels, and then extend the covariance descriptor to represent the global feature of scene images. In addition, a One-vs-Rest multiple classes SVM is employed to learning the combination classifiers of scene recognition in this paper, and the voting strategy, which combines multiple classifiers to increase classification accuracy[17], is used to mange the labels generated by classifiers.

The experiment results demonstrate the feasibility and validity of proposed scene recognition algorithm. Also, it is easy to implement the proposed algorithm due to the low-dimension of covariance descriptor.

Acknowledgment

The authors would like to thank the OpenCV and Libsvm library.

References

[1] L. Lu, K. Toyama and G.D. Hager, *A Two Level Approach for Scene Recognition*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR), 1, 688-695, Jun. 2005.

[2] N. Serrano, A.E. Savakis and J. Luo, *Improved Scene Classification using Efficient Low-level Features and Semantic Cues*, Pattern Recognition, 37(9), 1773-1784, 2004.

[3] A. Oliva and A. Torralba, *Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope*, International Journal of Computer Vision(IJCV), 42(3), 145-175, 2001.

[4] M. Szummer and R.W. Picard, *Indoor-outdoor image classification*, IEEE International Workshop Content-Based Access Image Video Database, 1998.

[5] A. Vailaya, M. Figueiredo, A.Jain and H.J. Zhang, *Image Classification for Content-based Indexing*, IEEE Transaction Image Processing, 10(1), 117-130, 2001.

[6] O. Tuzel, F. orikli and P. Meer, *Region Covariance: A Fast Descriptor for Detection and Classification*, Proceedings of the European Conference on Computer Vision, Graz, Austria, 2006.

[7] Wikipedia, $http://en.wikipedia.org/wiki/Image\_histogram$.

[8] R. Haralick, K. Shanmugam and I. Dinstein, *Texture features for image classification*, IEEE Transaction on System, Man and Cybernatic, 1973.

[9] S. Aksoy and R. Haralick, *Feature Normalization and Likelihood-Based Similarity Measures for Image Retrieval*, Pattern Recognition Letters, 22(5):563-582, 2001.

[10] S. Avidan and A. Shamir, *Seam Carving for content-Aware Image Resizing*, ACM Trasactions on Graphics, Volume 25, Number 3, SIGGRAPH 2007.

[11] B.E. Boser, I. Guyon and V. Vapnik *A Training Algorithm for Optimal Margin Classifiers*, Computational Learing Theory, 144-152, 1992.

[12] J.C. Burges, *A Tutorial on Support Vector Machines for Pattern Recognition*, Data Mining and Knowledge Discovery, 2:121-167, 1998.

[13] N. Cristianini and J.S. Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, ISBN:0521780195, Cambridge University Press, 2000.
[14] P. Bartlett and J. Shawe-Taylor *Generalization Performance of Support Vector Machines and Other Pattern Classifiers*, Advances in Kernel Methods - Support Vector Learning, MIT Press, Cambridge, USA, 1998.
[15] C. Cortes and V. Vapnik, *Support Vector Networks*, Machine Learning, 20:273-297, 1995.
[16] E.L. Allwein, R.E. Schapire and Y. Singer, *Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers*, Journal of Machine Learning Research, 2000.
[17] Y. Sun, M.S. Kamel and Andrew K.C. Wong, *Empirical Study on Weighted Voting Multiple Classifiers*, Pattern Recognition and Data Mining, 3686, 335-344, 2005.