Shape Priors and Discrete MRFs for Knowledge-based Segmentation

Ahmed Besbes[1,2], Nikos Komodakis[3], Georg Langs[4], and Nikos Paragios[1,2]
(1) Laboratoire MAS, Ecole Centrale Paris, Châtenay-Malabry, France
(2) Equipe GALEN, INRIA Saclay - Ile-de-France, Orsay, France
(3) Department of Computer Science, University of Crete, Heraklion, Greece
(4) CIR Lab, Department of Radiology, Medical University of Vienna, Austria
ahmed.besbes@ecp.fr, komod@csd.uoc.gr, georg.langs@meduniwien.ac.at, nikos.paragios@ecp.fr

Abstract

In this paper we introduce a new approach to knowledge-based segmentation. Our method consists of a novel representation to model shape variations as well as an efficient inference procedure to fit the model to new data. The considered shape model is similarity-invariant and refers to an incomplete graph that consists of intra and inter-cluster connections representing the inter-dependencies of control points. The clusters are determined according to the co-dependencies of the deformations of the control points within the training set. The connections between the components of a cluster represent the local structure while the connections between the clusters account for the global structure. The distributions of the normalized distances between the connected control points encode the prior model. During search, this model is used together with a discrete markov random field (MRF) based segmentation, where the unknown variables are the positions of the control points in the image domain. To encode the image support, a Voronoi decomposition of the domain is considered and regional based statistics are used. The resulting model is computationally efficient, can encode complex statistical models of shape variations and benefits from the image support of the entire spatial domain.

1. Introduction

Knowledge-based segmentation is a fundamental task of low and mid-level vision. In fields where a prior knowledge is available (like medical imaging), such a method carries on great potentials since the domain knowledge can be used to introduce constraints which improves the final reliability and accuracy of the segmentation result. In order to do so, one first has to determine a model representing these constraints and then an inference process which aims to combine the visual support with the prior knowledge.

The definition of the shape model involves a representation and a statistical model. Landmark-based representations are widely used in computer vision. Point-based representations [6] are a typical example of such methods that have been studied widely in the context of active contours and snakes [18] through continuous interpolation strategies. Implicit representations [14] are an alternative approach to model shapes that handle topological changes naturally, while being computationally inefficient. The above-mentioned methods reconstruct the shape through local or global interpolation.

Once the representation has been considered, the next task consists of modeling its variation within a training set in order to construct the prior. In this context, simple average models [3], principal component analysis [6], as well as their kernel variants [8], Gaussian densities [15], mixture models, and non-parametric priors were considered. These methods make an explicit assumption on the nature of the statistical behavior of the training set and then determine the optimal set of parameters towards representing the observed variations.

Image-based inference is the last issue to be addressed where one aims to combine the visual support with the prior model. To this end, a cost function that combines both edge-based as well as region-driven terms is often considered. The main challenge is to determine the corresponding optimal solution that is often challenging with gradient-based approaches [18]. On the other hand, discrete methods [2] could yield a better minimum of the objective function under certain constraints but the integration of global deformable priors [9, 16] is not straightforward. The aim of our approach is to address the above-mentioned limitations of conventional knowledge-based segmentation approaches.

To build a model that combines global and local deformable priors within the MRF framework, we consider a novel representation where control points are clustered according to their behavior within the training set using a

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

linear programming approach [11]. Clusters correspond to sets of points for which one can predict with certain confidence the positions given the position of the cluster center. This clustering is obtained through an inference process that measures the statistical similarity in terms of deformation between pairs of control points using shape maps [13]. On the other hand, the relative positions of cluster centers with respect to control points that do not belong to their clusters encode the global structure of the shape. Then, we model the shape variation through probability densities that encode the aforementioned local and global dependencies. The resulting framework can encode simple or complex distribution models according to the entropy of the observed system, unlike [10] where the model is assumed to be Gaussian. Such a model is represented using an incomplete graph having a structure derived from the training set, where the connections between the components of a cluster represent the local dependencies, while the connections between the clusters account for the global correlations between parts of the shape.

Then, inference consists of deforming the model consistently with the image information. The unknown parameters refer to the positions of the control points. In order, to determine the support from the image, we propose a Voronoi decomposition of the domain, defining a membership function that relates the pixels to the control points. The data term is then determined using this decomposition, while the prior term is expressed using the pairwise potentials between control points. Recent advances in the area of discrete optimization which explore the duality theorem of linear programing [12] are considered to recover the lowest potential of the objective function.

The remainder of this paper is organized as follows. In section 2 we introduce the shape model while in section 3 we propose the knowledge-based discrete segmentation framework. In section 4 we present the application and the experimental validation and the last section concludes the paper.

2. Shape Representation

Knowledge-based segmentation methods are based on the definition of a model which is then combined with image measurements towards object extraction. Classic approaches consist of representing the shape using a number of landmarks and learning their behavior using a training set.

2.1. A point-distribution model

Our shape model $\mathbf{S} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ consists of a set of n control points lying on its boundary. The contour of the shape can be recovered for example by interpolating the positions of these control points. The information carried by our model is described in a similarity-invariant manner, using the distances $\mathbf{d}_{ij}$ between pairs of control points $(\mathbf{x}_i, \mathbf{x}_j)$, normalized by the scale $\mathbf{d}$ of the object, or:

$$\mathbf{d}_{ij} = \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\mathbf{d}} \quad , \tag{1}$$

where $\mathbf{d} = \frac{2}{n.(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|$. Let us consider now a set $\mathcal{S} = \{s_1, \ldots, s_m\}$ of m instances of the object, where each example is represented using n control points, i.e. $s_u = \{x_1^u, \ldots, x_n^u\}, \forall u \in \{1, \ldots, m\}$. Hence, $\forall i \in \{1, \ldots, n\}$, the set $\mathcal{X}_i = \{x_i^1, ..., x_i^m\}$ represents instances of the i^{th} control point of the shape. In practice, this training set is obtained by manually labeling the landmarks for each instance of the shape, or by deducing the landmarks from the registration between a labeled shape and a set of non-labeled shapes.

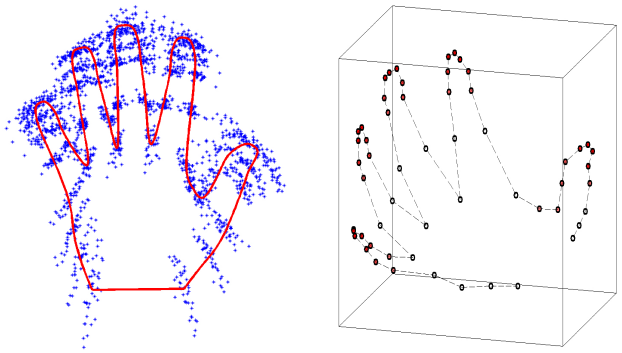
Then, given a statistical model, we learn from the training set the probability density distributions of the relative positions of the control points $p_{ij} \equiv p(\mathbf{x}_i, \mathbf{x}_j) \equiv p(\mathbf{d}_{ij})$. These $\frac{n.(n-1)}{2}$ densities enables us to describe the information contained in the training set. However, this representation suffers from redundancy and is expensive in terms of computation cost.

2.2. Removing redundancy

The task of eliminating the redundancy in the model, while preserving its ability to represent the data, is related to the minimum description length principle on one hand, and can be thought as a spectral clustering problem on the other hand. We aim to obtain as compact a model as possible assuming that the high dimensional data space can be approximated by a lower dimensional embedded manifold, which reduces the dimension of the problem significantly. Shape maps [13] handle precisely these two aspects, and are learned from the data in a way closely related to the diffusion maps [4], but using the compactness of models that describe sub sets of the entire data instead of the spatial distances or similarities between individual points. Therefore, we compute the shape map of the control points, using the training set $\{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$, and then we cluster the control points according to their mutual shape map distances. For a pair of control points $(\mathbf{x}_i, \mathbf{x}_j)$, the resulting map distance will be noted $\mathbf{d}_{sm}(\mathbf{x}_i, \mathbf{x}_j)$. A new clustering algorithm [11] was used for this final task, and is described in the section 2.3. The obtained clusters reflect the interdependencies between the control points, and refer to the parts of the object that have highly-correlated relative displacements.

2.3. Clustering via linear programming

Clustering refers to the process of organizing a set of objects into groups, where the members of each group are as similar to each other as possible. More formally, a common



(a) (b)

Figure 1: Model construction. (a) The considered training set. (b) The density or redundancy is color coded on the projection of the control points in the first 3 shape map dimensions. It reflects the coherence of local shape variation.

definition for clustering is the following one: suppose we are given a set of objects $\mathcal{V} = \{v_1, \ldots, v_n\}$ endowed with a distance function d that can measure the similarity between any two objects $v_i, v_j \in \mathcal{V}$. In such a case, the goal of clustering is to choose K objects, say, $\{c_1, c_2, \ldots, c_K\}$ from $\mathcal{V}$ (these will be referred to as the *clusters centers*), so that the obtained sum of distances between each object and its closest center is minimized, or:

$$\min_{c_1,c_2,\ldots,c_K} \sum_{v_i \in \mathcal{V}} \min_{c_k} d(v_i, c_k) \ . \tag{2}$$

A common drawback of many popular clustering techniques (such as the K-means algorithm) is that they need to be given the number K of clusters beforehand. However, this is problematic as this number is very often not known in advance. To address this issue, we will let this number vary as well and change the objective function of clustering so as to assign a penalty (denoted by $d(v_i, v_i)$) whenever an object v_i is chosen as a cluster center, or:

$$\min_{K,c_1,c_2,\ldots,c_K} \sum_{v_i \in \mathcal{V}} \min_{c_k} d(v_i, c_k) + \sum_{c_k} d(c_k, c_k) \ . \tag{3}$$

Another very bad symptom of many clustering techniques is that they are particularly sensitive to initialization. For instance, the K-means algorithm (which is one of the most commonly used clustering techniques) is doomed to fail if its initial cluster centers happen not to be near the actual cluster centers. To address this very important issue, we have used a novel clustering method. The main idea behind our method is to first formulate the clustering as a linear integer program as follows:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} d(v_i, v_j) x_{ij} \tag{4}$$

$$\text{s.t.} \sum_{j=1}^{n} x_{ij} = 1, \ \forall i \tag{5}$$

$$x_{ij} \leq x_{jj}, \ \forall i \neq j \tag{6}$$

$$x_{ij} \in \{0, 1\}, \ \forall i, j \tag{7}$$

In the above formulation, the binary variable x_{ij} (with $i \neq j$) indicates whether object v_i has been assigned to cluster center v_j or not, while the binary variable x_{jj} indicates whether object v_j has been chosen as a cluster center or not. It is then very easy to prove that the above linear integer program is actually equivalent to minimizing the objective function (3) of clustering. To this end, it suffices to observe that (5) simply expresses the fact that each object v_i can be assigned to exactly one cluster center v_j, while (6) simply expresses the fact that if any object v_i has been assigned to an object v_j, then v_j must be chosen as cluster center. To obtain an approximately optimal solution to the above integer program, we will then rely on first solving its linear programming relaxation and then "rounding" the relaxed solution in an appropriate manner. More details about the formulation of the problem and its optimization are given in [11]. In the validation section of [11], it is added that a constant penalty cost, roughly set to the median of the distances $d(v_i, v_j)$ is used in the experiments. We also considered the same penalty value for our tests.

In our case, the set of objects correspond to the control points $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, and the distance d corresponds to the aforementioned distance $\mathbf{d}_{sm}$ in section 2.2. We give in [Fig. 2(a)] an example of the output of this clustering process using the hand database [17].

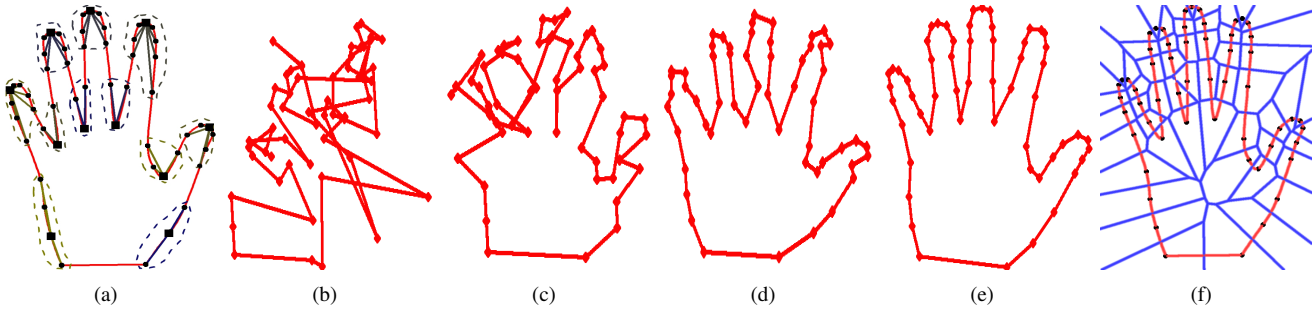
2.4. The shape model

Before proceeding, let us summarize the model we obtain after the clustering step. Our shape model $\mathbf{M} = (\mathbf{S}, \mathbf{P})$ is an incomplete graph. It consists of a set of control points (unary cliques) $\mathbf{S} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ lying on the boundary of the object, and a set $\mathbf{P} = \mathbf{P}_l \cup \mathbf{P}_g$ of pairs of control points (binary cliques), where $\mathbf{P}_l$ contains the "local" pairs, and $\mathbf{P}_g$ contains the "global" pairs, or:

$$(i,j) \in \mathbf{P}_l \iff \mathbf{x}_i \in C(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in C(\mathbf{x}_i) \tag{8}$$

$$(i,j) \in \mathbf{P}_g \iff \mathbf{x}_i \notin C(\mathbf{x}_j) \text{ or } \mathbf{x}_j \notin C(\mathbf{x}_i) \tag{9}$$

where $C(\mathbf{x}_i)$ is the cluster having $\mathbf{x}_i$ as center. Hence, each cluster center is connected to the control points in its cluster (local pairs) and to all the other control points (global pairs), which leads to a k-fan graph structure [7]. The novelty here consists in the method that defines automatically from the



(a) (b) (c) (d) (e) (f)

Figure 2: Model representation. (a) Control points clustered in 11 clusters: centers are represented by squares. (b)–(e) Deformation of a point cloud according to the shape prior term. (f) Voronoi decomposition of the model domain.

training data the number of fans and their centers. To each one of these pairs $(\mathbf{x}_i, \mathbf{x}_j)$ we associate a probability density distribution p_{ij} learned from the training set as previously stated in section 2.1. In practice, applying shape prior constraints to an initial set of random control points leads to an instance of the learned object, as showed in [Fig. 2(b)–2(e)]. The use of such prior in the segmentation framework is a much more interesting application and it is explored in the following.

3. Knowledge-based Segmentation

The main challenges of knowledge-based segmentation are: (i) appropriate modeling of shape variations, (ii) successful inference between the image and the manifold. Let us consider the simplest possible scenario that aims to detect an object of particular interest from the background in an image $\mathcal{I}$. We formulate the segmentation problem as an energy minimization problem. First, we introduce our model in the image at an initial position and state (it will be noted $\mathbf{M}^0$). Then, we search the optimal displacements $\vec{\mathbf{D}} = (\vec{\mathbf{d}}_1, \ldots, \vec{\mathbf{d}}_n)$ of the control points that give the best compromise between the pairwise prior constraints, encoded in our shape model $\mathbf{M}$, and the fidelity to the image information. Formally, the segmentation of the image $\mathcal{I}$ using the shape model $\mathbf{M}$ is given by:

$$\left(\vec{\mathbf{d}}_1, \ldots, \vec{\mathbf{d}}_n\right) = \underset{\vec{d}_i}{\operatorname{argmin}}\, E\left(\mathbf{M}^0, \mathcal{I}, \left(\vec{d}_1, \ldots, \vec{d}_n\right)\right) \quad . \tag{10}$$

The energy $E(\mathbf{M}^0, \mathcal{I}, \vec{D})$ of displacing the model in the image by the displacement vectors $\vec{D} = (\vec{d}_1, \ldots, \vec{d}_n)$ is the sum of a data-related term $V(\mathbf{S}^0 + \vec{D}, \mathcal{I})$ expressing the image cost of displacing the control points in $\mathbf{S}^0$ from their initial position, and a prior term $V(\mathbf{P}, \vec{D})$ expressing the cost of deforming the pairs in $\mathbf{P}$ from their initial position according to the displacement vectors and with respect to the prior learned distributions :

$$E\left(\mathbf{M}^0, \mathcal{I}, \vec{D}\right) = V\left(\mathbf{S}^0 + \vec{D}, \mathcal{I}\right) + V\left(\mathbf{P}, \vec{D}\right) \quad . \tag{11}$$

We explain in this section how we define these two energy terms, and we detail in particular the relationships between the control points and the image domain. We also develop an optimization procedure that enables to solve (10) using an efficient discrete optimization algorithm.

3.1. Regional Statistics & Image Segmentation

By applying the data-related cost, we seek the optimal separation of the object from the background in terms of visual properties. Let, p_{obj} and p_{bcg} be the conditional densities for these two hypotheses. Given that the control points $\mathbf{S}^0 + \vec{D} = \{\mathbf{x}_1^0 + \vec{d}_1, \ldots, \mathbf{x}_n^0 + \vec{d}_n\}$ form a closed boundary, they partition the image domain Ω into an object domain Ω_{obj} and a background domain Ω_{bcg}. To simplify the notation here, we will refer to a current configuration of the control points that will be noted $\mathbf{S}$. Then by considering the $-\log$ of the posterior probabilities, we express the cost $V(\mathbf{S}, \mathcal{I})$ using the regional statistics [21] as follows:

$$\begin{aligned} V(\mathbf{S}, \mathcal{I}) = & \sum_{y \in \Omega_{obj}} -\log\left(p_{obj}\left(\mathcal{I}\left(y\right)\right)\right) \\ & + \sum_{y \in \Omega_{bcg}} -\log\left(p_{bcg}\left(\mathcal{I}\left(y\right)\right)\right) \quad . \end{aligned} \tag{12}$$

In order to evaluate this component and associate it with the proposed shape representation, we decompose the image domain Ω according to the control points $\mathbf{S}$ by considering their Voronoi diagram [Fig. 2(f)]: $\Omega = \cup_{i=1}^{n} \Omega_i$, where Ω_i is the Voronoi cell associated with the control point $\mathbf{x}_i$. By intersecting these Voronoi cells with the object domain and the background domain, we obtain the partition $\Omega = \cup_{i=1}^{n} (\Omega_{obj,i} \cup \Omega_{bcg,i})$ that relates each pixel of the image to one control point, and specifies its class. Then, one can decompose the global image term (12) into sub-terms

which are defined at the partition cells as follows:

$$V(\mathbf{S},\mathcal{I}) = \sum_{i=1}^{n} V_i(\mathbf{x}_i,\mathcal{I}) \tag{13}$$

$$\begin{aligned}\text{with } V_i(\mathbf{x}_i,\mathcal{I}) = &\sum_{y\in\Omega_{obj,i}} -\log\left(p_{obj}\left(\mathcal{I}\left(y\right)\right)\right) \\ &+ \sum_{y\in\Omega_{bcg,i}} -\log\left(p_{bcg}\left(\mathcal{I}\left(y\right)\right)\right)\end{aligned} \tag{14}$$

being the image-related cost associated with the control point position $\mathbf{x}_i$. These terms can be calculated very efficiently per class by combining rasterization techniques and fast integral computing over polygons [19]. We should note that this term uses the entire image domain to determine the image support and can be replaced either using more complex descriptors, or through edge-driven support. In practice, we used simple Gaussian models and mixture of Gaussians models for the object and the background. Such a component will perform well if the data support is strong but will fail to deal with noise, clutter, missing parts, etc. The use of prior knowledge on the expected geometry of the shape could address the above mentioned limitations.

3.2. Prior Knowledge & Image Segmentation

In the context of our approach, we have defined the shape model as an incomplete graph. Furthermore, we were able to determine an approximate density of this model using a small number of joint densities. In order to impose the prior, we minimize the cost $V(\mathbf{P},\vec{D})$ that we decompose over all the pairs:

$$\begin{aligned}V(\mathbf{P},\vec{D}) = \alpha \underbrace{\sum_{(i,j)\in\mathbf{P}_l} V_{ij}(\mathbf{x}_i^0+\vec{d}_i,\mathbf{x}_j^0+\vec{d}_j)}_{\text{local prior cost}} \\ + \beta \underbrace{\sum_{(i,j)\in\mathbf{P}_g} V_{ij}(\mathbf{x}_i^0+\vec{d}_i,\mathbf{x}_j^0+\vec{d}_j)}_{\text{global prior cost}} \ ,\end{aligned} \tag{15}$$

$$\text{with } V_{ij}(\mathbf{x}_i^0+\vec{d}_i,\mathbf{x}_j^0+\vec{d}_j) = -\log\left(p_{ij}(\mathbf{x}_i^0+\vec{d}_i,\mathbf{x}_j^0+\vec{d}_j)\right) \ . \tag{16}$$

This model allows for the encoding of global dependencies as local combinations of individual pairwise densities. Some examples of the impact of this term for a random collection of points with respect to the hand model [Fig. 2(a)] are shown in [Fig. 2(b)-2(e)]. The parameters (α,β) control the relative influence of inter and intra cluster constraints.

One can now integrate the data term with the prior term towards knowledge-based segmentation, by combining (11), (13) and (15):

$$\begin{aligned}E\left(\mathbf{M}^0,\mathcal{I},\vec{D}\right) = &\sum_{i=1}^{n} V_i(\mathbf{x}_i^0+\vec{d}_i,\mathcal{I}) \\ &+ \alpha \sum_{(i,j)\in\mathbf{P}_l} V_{ij}(\mathbf{x}_i^0+\vec{d}_i,\mathbf{x}_j^0+\vec{d}_j) \\ &+ \beta \sum_{(i,j)\in\mathbf{P}_g} V_{ij}(\mathbf{x}_i^0+\vec{d}_i,\mathbf{x}_j^0+\vec{d}_j) \ .\end{aligned} \tag{17}$$

3.3. The energy minimization

The optimization of this cost function (17) in the continuous domain is rather problematic. One can expect that it is not convex and therefore a gradient-based optimization will fail. In order to optimize such a cost function, we consider recent results from discrete optimization.

We make two assumptions that are most often verified in practice. First, the initial positions of the control points are within the image domain, and that is why we can assume an upper bound on the maximum displacements that would lead to the solution. Second, we consider that the precision required about the solution is specified, which enables to choose a quantization step of the displacement vectors $\vec{D}$. Then, we can approximate the continuous deformations of our shape model towards the solution by a finite set of displacements vectors $\vec{D} = \{\vec{d^1},\ldots,\vec{d^z}\}$. Let $\mathcal{L} = \{1,\ldots,z\}$ be the set of labels associated the quantization $\vec{D} = \{\vec{d^1},\ldots,\vec{d^z}\}$ of the displacements. Then, displacing the control point $\mathbf{x}_i^0$ by the vector $\vec{d^{l_i}}$ is equivalent to assigning the label l_i to $\mathbf{x}_i^0$, and the minimization of the energy in (17) can be written as a labeling problem, or:

$$(\mathbf{l}_1,\ldots,\mathbf{l}_n) = \underset{l_i\in\mathcal{L}}{\operatorname{argmin}}\ E\left(\mathbf{M}^0,\mathcal{I},(l_1,\ldots,l_n)\right) \ , \tag{18}$$

$$\begin{aligned}\text{with } E\left(\mathbf{M}^0,\mathcal{I},(l_1,\ldots,l_n)\right) = &\sum_{i=1}^{n} V_i(\mathbf{x}_i^0,l_i) \\ &+ \alpha \sum_{(i,j)\in\mathbf{P}_l} V_{ij}(\mathbf{x}_i^0,\mathbf{x}_j^0,l_i,l_j) \\ &+ \beta \sum_{(i,j)\in\mathbf{P}_g} V_{ij}(\mathbf{x}_i^0,\mathbf{x}_j^0,l_i,l_j)\end{aligned} \tag{19}$$

where $V_i(\mathbf{x}_i^0,l_i) = V_i(\mathbf{x}_i^0+\vec{d}_i,\mathcal{I})$ and $V_{ij}(\mathbf{x}_i^0,\mathbf{x}_j^0,l_i,l_j) = V_{ij}(\mathbf{x}_i^0+\vec{d}_i,\mathbf{x}_j^0+\vec{d}_j)$. In such a context, the problem of finding the most appropriate deformation of the initial shape can be expressed using an MRF with singleton and pairwise interactions between the control points. We should note that such an approach is invariant to translation, rotation and scale (due to the definition of (1)). Recovering the optimal solution of this objective function is known to be an NP-hard problem and the complexity is influenced mostly

from the pairwise potentials function. Hence, we consider an approximate solution to the labeling problem using the Primal-Dual algorithm [12].

The cardinality of the label set is quite important since on one hand it defines the accuracy of the search, while on the other hand increases the complexity of the algorithm. In order to address the above mentioned issues, first we consider an approach that is incremental in terms of displacements while reducing the number of interactions between the nodes of the graph, and retaining the ability to encode the global structure. To this end, we cope with the accuracy issue, that is closely related to the quantization of $\vec{D}$, by using a pyramidal coarse-to-fine approach. Each level of the pyramid corresponds to a quantization step that is refined in the following level. To speed up the convergence in each level of the pyramid, we also adopt an incremental approach in terms of the label set, where in each iteration t we look for the set of labels that will improve the current solution by minimizing:

$$E^t(l_1,\ldots,l_n) = \sum_{i=1}^{n} V_i(\mathbf{x}_i^{t-1}, l_i) + \alpha \sum_{(i,j)\in\mathbf{P}_l} V_{ij}(\mathbf{x}_i^{t-1}, \mathbf{x}_j^{t-1}, l_i, l_j) + \beta \sum_{(i,j)\in\mathbf{P}_g} V_{ij}(\mathbf{x}_i^{t-1}, \mathbf{x}_j^{t-1}, l_i, l_j) \quad , \tag{20}$$

$$\text{with } \mathbf{x}_i^t = \mathbf{x}_i^0 + \sum_{\tau=1}^{t} \vec{d}^{\,\mathbf{l}_i(\tau)} \quad , \tag{21}$$

with $\mathbf{l}_i(\tau)$ being the optimal label associated with the i^{th} control point at time τ. Towards computational efficiency and localization of a good minimum, we adopt a fast and efficient method for the optimization, the Primal-Dual algorithm [12] that is based on linear programing and takes benefit of the duality theorem. The main challenge of optimizing the above objective function relates with the fact the we have arbitrary pair-wise potentials. Therefore the use of method like graph-cuts [1] is prohibited while at the same time the use of more advanced optimization like belief-propagation networks [20] is also problematic due to the structure of the graph.

4. Experimental Validation

In order to validate the performance of our method, we considered the application of modeling the hand using a 2D 40-example dataset of annotated left hands, showing different relative finger positions, hand sizes, and texture [17]. On each hand contour, 56 landmarks were used to describe the structure. We performed clustering in the distribution space as described in section 2.3, using shape maps [13]. The clustering provided 11 clusters shown in [Fig. 2(a)]. The constructed model was used as a shape constraint as shown in [Fig. 2(b)-2(e)], and applied in different segmentation settings. We considered a multi-scale implementation of the approach using gradually an increasing number of control points to accelerate convergence. First, we segmented correctly 37 out of the 40 examples of the database. Examples of the results we obtained are shown in [Fig. 4(a)]. We also compared quantitatively our method to AAM segmentation [Fig. 3]. We can see in these boxplots that our algorithm performs better quantitatively with examples where the forearm is hidden by a sleeve. In the case of nude forearms, the data term drives the model to "oversegment" the hand in comparison with the ground truth, which explains our results. These "oversegmentations" are visually correct (especially the fingers are correctly segmented) as we can see in [Fig. 4(a)]. The three examples where our method did not succeed are particularly difficult because they exhibit occlusion between the fingers, which can cause folding in the evolving contour. Towards checking the robustness of the method, we removed some hands parts for several examples, and despite the important missing structure, the results were quite satisfactory as shown in [Fig. 4(b)]. The prior weight in these cases was increased, and enforced the correct segmentation, as the data term was less reliable. Furthermore, to validate the robustness of our method, we added severe Gaussian noise to the database images. The segmentations obtained in [Fig. 4(a)] are completely or almost recovered, thanks to the prior knowledge, as it is shown in [Fig. 4(c)]. Eventually, we used our segmentation method in a real setting, on hand video frames, with a cluttered background and partial occlusion cases. [Fig. 4(d)] gives some examples of the obtained segmentations. We could reproduce the result we obtained on the noisy images using AAM segmentation [5], but this algorithm could not reproduce our results for the occlusion cases .In our experiments, one iteration lasts approximately 1s using a non-optimized program, on a DELL Duo Computer (3GHz,3GB).

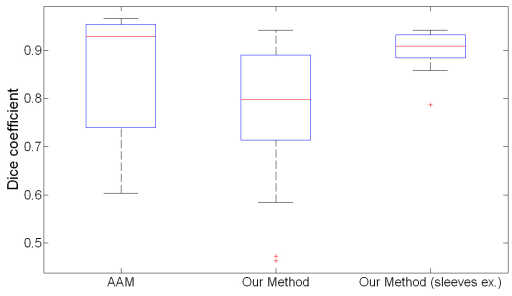


Figure 3: Boxplots of dice overlap coefficients comparing our method to AAM.

5. Discussion

In this paper, we have proposed a novel approach to knowledge-based segmentation. Our main contribution consists of modeling the co-dependencies between control points deformations towards a compact, sparse but efficient shape representation using an incomplete graph that was determined through an unsupervised clustering approach on the relevance of statistical behavior of control points deformations. This representation is combined with a data term like regional statistics in order to perform inference of the model location or analogously segmentation in new image data. To this end, a MRF is considered where singleton potentials account for the image support while pairwise potentials encode the shape prior. Our approach can claim certain optimality properties thanks to the efficient linear programming optimization techniques considered in this paper. Furthermore, our approach makes full use of the regional statistics and the obtained minimum is the one corresponding to the entire image potential. Our framework is however general, and we can imagine the use of different data terms that are more suitable to other applications or in different settings.

The proposed approach aims to optimize the connectivity of the graph nodes, and learns the structure and the local deformation statistics of an object from a set of training examples. In terms of clustering, the distance between the observations has a critical impact and should be further investigated. The temporal aspect of priors is also of great importance with potential applications to motion analysis, scene understanding and medical imaging. Therefore, extending the current framework in this direction would be also investigated.

Acknowledgments

We thank Radhouène Neji for the insightful remarks on the graph construction and for his interesting suggestions.

References

[1] Y. Boykov and G. Funka-Lea. Graph cuts and efficient n-d image segmentation. *Int. J. Comput. Vision (IJCV)*, 70(2):109–131, 2006.

[2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE T. Pattern Anal. (PAMI)*, 23(11):1222–1239, 2001.

[3] Y. Chen, F. Huang, H. D. Tagare, M. Rao, D. Wilson, and E. A. Geiser. Using prior shape and intensity profile in medical image segmentation. In *ICCV '03: Proc. 9th Int. Conf. Comput. Vision*, pages 1117–1124, 2003.

[4] R. R. Coifman and S. Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal. (ACHA)*, 21(1):5–30, 2006.

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE T. Pattern Anal. (PAMI)*, 23(6):681–685, 2001.

[6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Comput. Vis. Image Und. (CVIU)*, 61(1):38–59, 1995.

[7] D. Crandall, P. Felzenszwalb, and D. Huttenlocher. Spatial priors for part-based recognition using statistical models. In *CVPR '05: Proc. 2005 Conf. Comput. Vision & Pattern Recogn.*, pages 10–17, 2005.

[8] D. Cremers, T. Kohlberger, and C. Schnörr. Nonlinear shape statistics in Mumford–Shah based segmentation. In *ECCV '02: Proc. 7th Eur. Conf. Comput. Vision*, pages 93–108, 2002.

[9] D. Freedman and T. Zhang. Interactive graph cut based segmentation with shape priors. In *CVPR '05: Proc. 2005 Conf. Comput. Vision & Pattern Recogn.*, pages 755–762, 2005.

[10] L. Gu, E. Xing, and T. Kanade. Learning gmrf structures for spatial priors. In *CVPR '07: Proc. 2007 Conf. Comput. Vision & Pattern Recogn.*, pages 1–6, 2007.

[11] N. Komodakis, N. Paragios, and G. Tziritas. Clustering via lp-based stabilities. In *NIPS '08: Adv. Neural Inf. Proc. Syst. 21*, 2008.

[12] N. Komodakis, G. Tziritas, and N. Paragios. Performance vs computational efficiency for optimizing single and dynamic mrfs: Setting the state of the art with primal dual strategies. *Comput. Vis. Image Und. (CVIU)*, 112(1):14–29, 2008.

[13] G. Langs and N. Paragios. Modeling the structure of multivariate manifolds: Shape maps. In *CVPR '08: Proc. 2008 Conf. Comput. Vision & Pattern Recogn.*, 2008.

[14] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys.*, 79:12–49, 1988.

[15] M. Rousson and N. Paragios. Prior knowledge, level set representations & visual grouping. *Int. J. Comput. Vision (IJCV)*, 76(3):231–243, 2008.

[16] T. Schoenemann and D. Cremers. Globally optimal image segmentation with an elastic shape prior. In *ICCV '07: Proc. 11th Int. Conf. Comput. Vision*, pages 1–6, 2007.

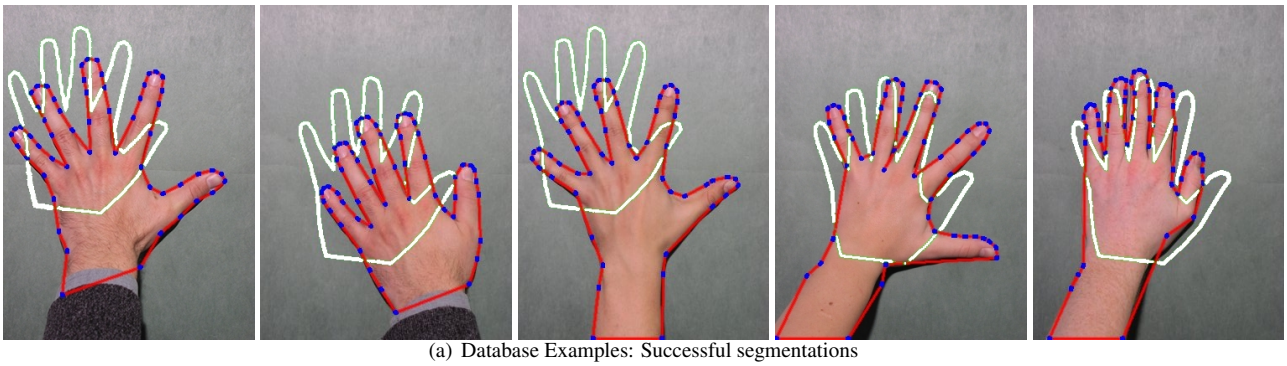
[17] M. B. Stegmann and D. D. Gomez. A brief introduction to statistical shape analysis, 2002.

[18] D. Terzopoulos, A. Witkin, and M. Kass. Constraints on deformable models: recovering 3d shape and nongrid motion. *Artif. Intell. (AI)*, 36(1):91–123, 1988.

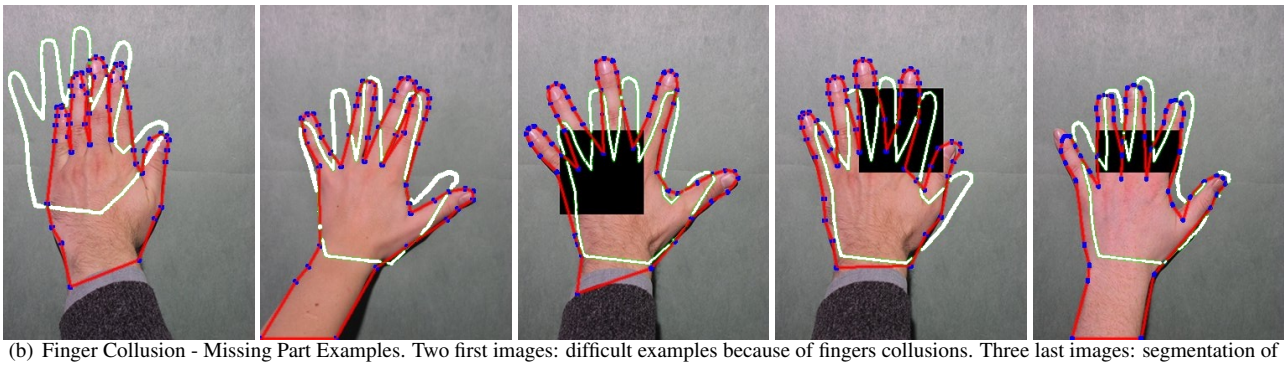
[19] G. B. Unal, A. J. Yezzi, and H. Krim. Information-theoretic active polygons for unsupervised texture segmentation. *Int. J. Comput. Vision (IJCV)*, 62(3):199–220, 2004.

[20] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE T. Inform. Theory (IT)*, 51(7):2282–2312, 2005.

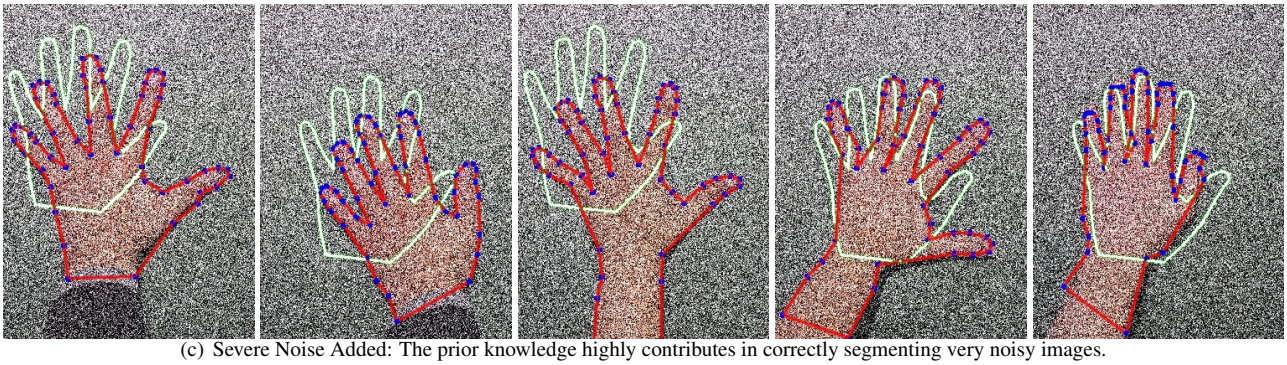
[21] S. C. Zhu and A. L. Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE T. Pattern Anal. (PAMI)*, 18(9):884–900, 1996.



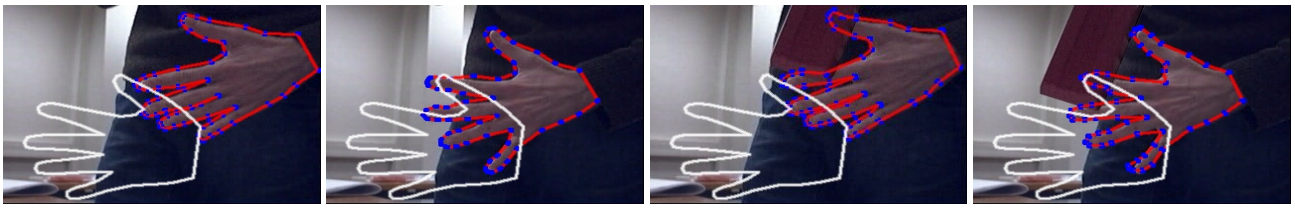
(a) Database Examples: Successful segmentations



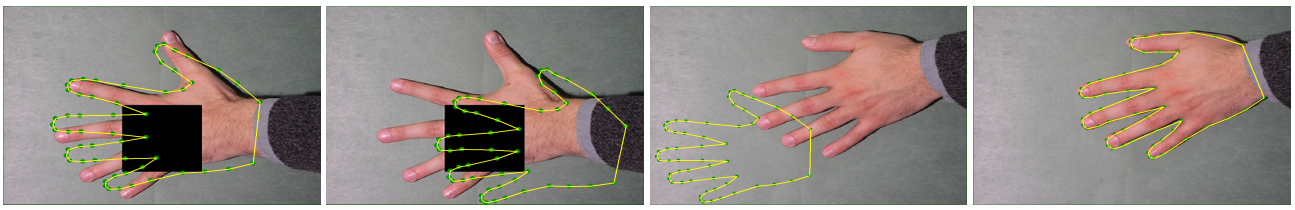
(b) Finger Collusion - Missing Part Examples. Two first images: difficult examples because of fingers collusions. Three last images: segmentation of hands with missing parts.



(c) Severe Noise Added: The prior knowledge highly contributes in correctly segmenting very noisy images.



(d) Video Frames - Partial Occlusions. Real video frames: cluttered background and occlusions.



(e) AAM results: succeeds with the learning examples but fails with occlusions. Initialization on the left - result on the right.

Figure 4: Model-based segmentation of the hand. Initialization is shown in white, segmentation in red, and the final control points positions in blue.