HAL
open science

Pose Invariant Deformable Shape Priors Using L1 Higher Order Sparse Graphs

Bo Xiang, Nikos Komodakis, Nikos Paragios

To cite this version:

Bo Xiang, Nikos Komodakis, Nikos Paragios. Pose Invariant Deformable Shape Priors Using L1 Higher Order Sparse Graphs. 9th International Symposium on Visual Computing - ISVC 2013, Jul 2013, Rethymnon, Greece. hal-00856978

HAL Id: hal-00856978

https://inria.hal.science/hal-00856978

Submitted on 2 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pose Invariant Deformable Shape Priors Using L_1 Higher Order Sparse Graphs

Bo Xiang[1,2], Nikos Komodakis[1,3,4], and Nikos Paragios[1,2]

[1] Center for Learning and Visual Computing,
Ecole Centrale de Paris - Ecole des Ponts ParisTech, France
[2] Equipe GALEN, INRIA Saclay, Île-de-France, France
[3] Université Paris-Est, Ecole des Ponts ParisTech, France
[4] UMR Laboratoire d'Informatique Gaspard-Monge, CNRS, France

Abstract. In this paper we propose a novel method for knowledge-based segmentation. We adopt a point distribution graphical model formulation which encodes pose invariant shape priors through L_1 sparse higher order cliques. Local shape deformation properties of the model can be captured and learned in an optimal manner from a training set using dual decomposition. These higher order shape terms are combined with conventional visual ones aiming at maximizing the posterior segmentation likelihood. The considered graphical model is optimized using dual decomposition and is used towards 2D (computer vision) and 3D object segmentation (medical imaging) with promising results.

1 Introduction

Knowledge-based segmentation is a fundamental problem in computer vision and medical imaging. The central idea is to combine the image information with prior knowledge learned from examples (mostly regarding the geometric properties of the class of objects) in order to cope with occlusions, non-discriminative visual support and noise.

Early approaches adopted snake-based formulations and sought to impose constraints on the interpolation coefficients of the basis functions [1]. Active shape models [2] and their visual variance have been a fundamental step towards modeling globally shape variations through principal component analysis on a set of training examples and use of the associated sub-space for manifold-constrained segmentation. Level set methods have been also endowed with priors either including simple average models [3], subspaces [4] or to certain extend pose invariance [5].

The graph-theoretic approaches were also considered in knowledge-based segmentation. In [6], shape constraints were used iteratively to modify the graph potentials towards imposing prior knowledge by the means of mean shape. Direct modeling of prior knowledge within graphs have been presented either using global priors within the random walker algorithm [7] or through modeling of the segmentation over the optimization of a graph corresponding to the point distribution model. For example, prior knowledge was modeled through statistical definition of the pair-wise constraints in the works of [8, 9]. Unfortunately these methods were not pose invariant (*i.e.* invariant to translation, rotation and scale of the global shape) and could not model properly data

support. This problem was partially addressed in [10] through a fully connected complex graph with computational complexity being the main bottleneck.

In this paper, we propose a novel method to encode pose invariant shape priors through L_1 sparse higher order graphs. We adopt a point distribution model that involves pair-wise and higher order cliques. Pair-wise terms are used to account for data support, while second order potentials encode the local shape deformation statistics. The subset of cliques from all possible second order cliques is learned through dual decomposition. This is to provide the best possible reconstruction of the observed shape variation, while being as compact as possible. This model is applied to the image and combined with visual information (edges, regional statistics) towards knowledge-based segmentation. Hand-pose segmentation and 3D left ventricle segmentation are used as examples to demonstrate the potential of the method.

The remaining of the paper is organized as follows. Section 2 presents the considered shape model and the L_1 sparse prior, while section 3 integrates this prior to a visual segmentation model. Implementation details and experimental validation are part of section 4, while the last section concludes the paper.

2 L_1 Sparse Higher Order Graph Shape Representations

Shape prior modeling is the fundamental task in knowledge-based segmentation. Based on a shape representation, a statistical shape model is built from a training set in order to have: (1) the ability to describe the shape variation of the object of interest, (2) a compact representation of the shape constraints and (3) the facility to be encoded in an inference process towards image segmentation.

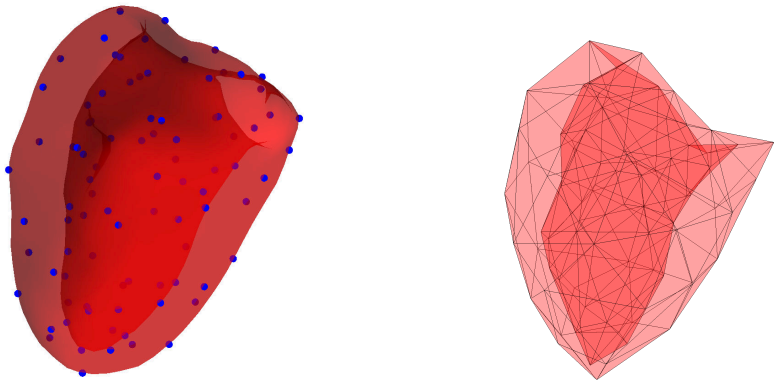
2.1 Shape representation

We use a point-based model $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ to represent the shape. It consists of a set $\mathcal{V} = \{1, \cdots, n\}$ of n control points distributed on the boundary of the object of interest (*e.g.* see Fig.1 (a)), where $\mathbf{x}_{i\in\mathcal{V}}$ denotes the coordinates of the i-th point. Additionally, the local interactions of the shape model is represented by cliques, where each clique is a subset of the point set $\mathcal{V}$. Considering the size of a clique being three, we denote a clique set $\mathcal{C} = \{(i,j,k)|i,j,k \in \mathcal{V} \text{ and } i \neq j \neq k\}$ consisting of all possible combinations of three points.

For a triplet clique $c = (i,j,k) \in \mathcal{C}$, the geometric shape of the clique $\mathbf{x}_c = (\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$ can be defined by its two inner angles (α_c, β_c) which are *pose invariant*, *i.e.* invariant to translation, rotation and scale of the shape.

$$\alpha_c = \cos^{-1} \frac{\overrightarrow{\mathbf{x}_i\mathbf{x}_j} \cdot \overrightarrow{\mathbf{x}_i\mathbf{x}_k}}{\|\mathbf{x}_i\mathbf{x}_j\| \, \|\mathbf{x}_i\mathbf{x}_k\|}, \;\; \beta_c = \cos^{-1} \frac{\overrightarrow{\mathbf{x}_j\mathbf{x}_k} \cdot \overrightarrow{\mathbf{x}_j\mathbf{x}_i}}{\|\mathbf{x}_j\mathbf{x}_k\| \, \|\mathbf{x}_j\mathbf{x}_i\|} \tag{1}$$

Given a training set of K shape instances $\{\mathbf{X}^k\}_{k=1}^K$, we assume that point correspondences exist between the point distribution models within the training set, without assuming that shapes have been brought to the same referential. Using a standard probabilistic model (*e.g.* Gaussian Distributions), the probability distributions $p_c(\alpha_c, \beta_c)$ of clique c are learned from K instances $\{(\alpha_c^k, \beta_c^k)\}_{k=1}^K$.



(a) (b)

Fig. 1. 3D myocardium shape model. (a) Control points (blue dots). (b) Triangulated mesh.

Assuming the independence between cliques, the probability $p(\mathbf{X})$ of a shape configuration $\mathbf{X}$ can be formulated by the accumulation of all triplet clique constraints using the probability distribution $p_c(\alpha_c, \beta_c)$.

$$p(\mathbf{X}) \propto \prod_{c \in \mathcal{C}} p_c(\alpha_c, \beta_c) \tag{2}$$

In this manner, our shape prior model inherits pose invariance so that neither training samples nor testing shapes need to be aligned in a common coordinates frame. This model using local statistics can also capture shape variations even with a small number of training examples. However, it is extremely complex due to the excessive number of higher order cliques. Furthermore, assuming independence between cliques is an invalid assumption since there can be strong correlation between them, at least at local scale. Last but not least, the significance of the different triplets, towards capturing the observed deformations of the training set, is not the same.

2.2 L_1 higher order MRF learning via dual decomposition

To address the above issues, we work on a compact representation of shape priors while preserving its ability to describe shape variability. First of all, we cast the above shape prior modeling as a higher-order Markov Random Fields (MRF) optimization problem, where we search for an optimal shape configuration $\mathbf{X}^{\text{opt}}$.

$$\mathbf{X}^{\text{opt}} = \arg\min_{\mathbf{X}} E(\mathbf{X}) \tag{3}$$

Given a graph $G = (\mathcal{V}, \mathcal{C})$ consisting of a node set $\mathcal{V}$ and a clique set $\mathcal{C}$, we associate a control point to a node and a triplet of control points to a clique. Let $\mathbf{x}_{i \in \mathcal{V}}$ denote the latent variable (*i.e.* the coordinates of a point) of node i. The MRF energy $E(\mathbf{X})$ is then

defined by higher-order (triplet) potentials h_c.

$$E(\mathbf{X}) = \sum_{c \in \mathcal{C}} w_c h_c(\mathbf{x}_c) \tag{4}$$

where $h_c(\mathbf{x}_c) = -\log p_c(\alpha_c, \beta_c)$. We point out that the above energy is parameterized[1] by introducing an additional vector of parameters $\mathbf{w} = \{w_{c \in \mathcal{C}}\}$ containing one component w_c per clique c. The role of the introduced vector $\mathbf{w}$ is essentially to give a different weight on the contribution of each clique in order to select cliques are going to be retained in the shape prior model. For instance, a clique c is ignored if the corresponding element is zero, *i.e.* if $w_c = 0$.

To estimate this vector $\mathbf{w}$, we use a MRF training procedure during which we impose a sparsity-enforcing prior on the vector $\mathbf{w}$ in order to eliminate as many redundant cliques as possible. Let $\{\mathbf{X}^k\}_{k=1}^K$ be the training set of shape instances. A max-margin learning formulation is employed for computing the vector $\mathbf{w}$, in which case we must minimize the following regularized empirical loss:

$$\min_{\mathbf{w}} \lambda ||\mathbf{w}||_1 + \sum_{k=1}^{K} \mathrm{Loss}_G(\mathbf{X}^k; \mathbf{w}) \tag{5}$$

In the above expression, the term $\lambda ||\mathbf{w}||_1$ is a sparsity inducing L_1-norm regularizer, and the term $\mathrm{Loss}_G(\mathbf{X}^k; \mathbf{w})$ denotes the hinge-loss with respect to $\mathbf{X}^k$ for the MRF defined on the graph G.

$$\mathrm{Loss}_G(\mathbf{X}^k; \mathbf{w}) = E(\mathbf{X}^k; \mathbf{w}) - \min_{\mathbf{X}} \big(E(\mathbf{X}; \mathbf{w}) - \Delta(\mathbf{X}, \mathbf{X}^k)\big) \tag{6}$$

where $\Delta(\mathbf{X}, \mathbf{X}')$ represents a dissimilarity measure between two solutions $\mathbf{X}$ and $\mathbf{X}'$. Intuitively, the above hinge-loss (6) expresses the fact that we should ideally adjust $\mathbf{w}$, such that the energy of the ground truth shape $E(\mathbf{X}^k; \mathbf{w})$ should be smaller than the energy of any other shape $E(\mathbf{X}; \mathbf{w})$ by at least a margin specified by the dissimilarity function $\Delta(\mathbf{X}, \mathbf{X}^k)$.

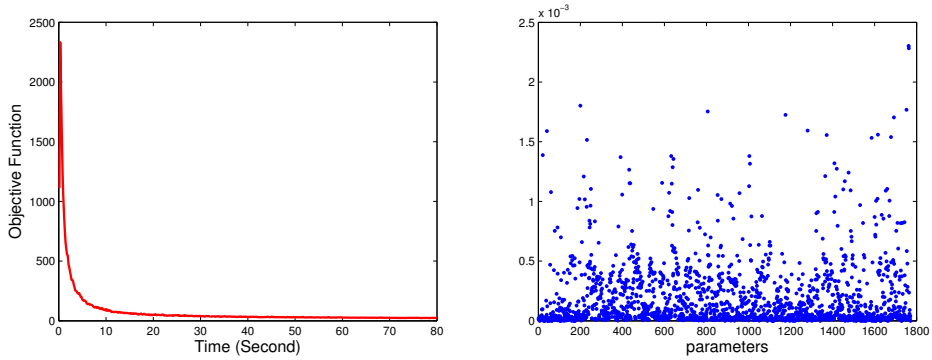
There are two main challenges that we need to deal with in this case: (i) The MRF $E(\mathbf{X}; \mathbf{w})$ (4) that we want to train contains high-order terms, (ii) The learning must take account of the fact that, if $\mathbf{X}^k$ is a ground truth shape, then any transformed shape instance $T(\mathbf{X}^k)$ under a similarity transformation T is an equally good solution and should not be penalized during training, *i.e.* it should hold $\Delta(T(\mathbf{X}), \mathbf{X}) = 0$.

In order to deal with (ii), we choose our dissimilarity function $\Delta(\mathbf{X}, \mathbf{X}')$ that decomposes into the following higher-order terms.

$$\Delta(\mathbf{X}, \mathbf{X}') = \sum_{c \in \mathcal{C}} \delta_c(\mathbf{x}_c, \mathbf{x}'_c) \tag{7}$$

where the term $\delta_c(\mathbf{x}_c, \mathbf{x}'_c)$ equals 0 if two triplets of points $\mathbf{x}_c$ and $\mathbf{x}'_c$ are *similar*, otherwise it equals 1. The similarity property of triplets can be defined using the angle

[1] With a slight abuse of notation, symbols $E(\mathbf{X})$ and $E(\mathbf{X}; \mathbf{w})$ will hereafter be used interchangeably for denoting the energy of an MRF.



(a) (b)

Fig. 2. MRF learning with hand dataset. (a) Primal objective function. (b) Learned parameters $\mathbf{w}$.

representation (1) which is invariant to similarity transformation, *i.e.* if $\alpha_c = \alpha'_c$ and $\beta_c = \beta'_c$, then $\mathbf{x}_c$ and $\mathbf{x}'_c$ are similar.

As a result of the above, not only the MRF energy $E(\mathbf{X}; \mathbf{w})$ but also the dissimilarity function $\Delta(\mathbf{X}, \mathbf{X}')$ contain high-order terms in our case. To deal with this challenge, we make use of the recently proposed dual decomposition framework for MRF learning [11] that can efficiently handle the training of high-order models. Such a framework essentially manages to reduce the task of training a complex high-order model on the graph G (*i.e.* minimizing the regularized empirical loss (5)) to the much easier task of training in parallel a series of slave MRFs defined on subgraphs.

The only restrictions that must be obeyed by these subgraphs are that (i) their union should cover the original graph G, and (ii) one should be able to minimize the energy of the so-called loss-augmented slave MRFs defined on these subgraphs. In our case, we choose one subgraph corresponding to each clique $c = \{i, j, k\} \in \mathcal{C}$ of graph G, in which case the loss-augmented energy of the resulting slave MRF is given by:

$$E_c(\mathbf{x}_c; \mathbf{w}) = \theta_i(\mathbf{x}_i) + \theta_j(\mathbf{x}_j) + \theta_k(\mathbf{x}_k) + w_c h_c(\mathbf{x}_c) - \delta(\mathbf{x_c}, \mathbf{x}_c^k) \tag{8}$$

where $\theta_i, \theta_j, \theta_k$ denote the unary potentials of the slave MRF. Such an energy is indeed possible to be optimized in our case (since we are using a discrete label set), thus leading to a very efficient stochastic subgradient learning scheme based on dual-decomposition.

To this end, we can achieve a proper vector $\mathbf{w}$ (*e.g.* see Fig.2(b)) from the MRF training process. Due to the sparsity regularizer, a large number of the cliques will be endowed with zero-value weight, *i.e.* $w_c = 0$. By eliminating these cliques, we obtain a sparse structure to model the shape prior. In practice, the sparse graph $\mathcal{G} = (\mathcal{V}, \mathcal{F})$ is composed by the cliques whose corresponding weights are above the threshold t, *i.e.* $\mathcal{F} = \{c | w_c > t, c \in \mathcal{C}\}$, while the size of the graph $\mathcal{G}$ is much smaller than the complete graph G, *i.e.* $|\mathcal{F}| << |\mathcal{C}|$.

3 Knowledge-based Segmentation

Now we integrate our compact shape model into knowledge-based image segmentation, aiming to recover the optimal instance of the learned manifold in an observed image. We formulate the segmentation problem within a higher-order MRF modeling framework.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{D})$ denote a hypergraph with a node set $\mathcal{V}$ and a clique set $\mathcal{D}$. The node set $\mathcal{V}$ is associated with the point-based model. The clique set $\mathcal{D} = \mathcal{E} \cup \mathcal{F}$ consists of two types of cliques: (1) The clique set $\mathcal{E}$ determines the boundary of the shape. It consists of the pairs of points (line segments on the closed curve) in 2D cases or the triplets of points (triangulated faces on the mesh, *e.g.* see Fig.1(b)) in 3D cases, (2) The clique set $\mathcal{F}$ represents the local interactions of the shape model using triplet cliques.

Let $\mathbf{x}_{i\in\mathcal{V}}$ denote the latent variable (*i.e.* the coordinates of point i), and $\mathbf{X} = (\mathbf{x}_i)_{i\in\mathcal{V}}$ be a configuration of all the node variables. The segmentation problem is formulated as an energy minimization, estimating the optimal point positions.

$$\begin{aligned} \mathbf{X}^{\text{opt}} &= \arg\min_{\mathbf{X}} E(\mathbf{X}) \\ E(\mathbf{X}) &= E_{\text{data}}(\mathbf{X}) + E_{\text{prior}}(\mathbf{X}) \end{aligned} \tag{9}$$

while the MRF energy $E(\mathbf{X})$ encodes both visual support (defined on the clique set $\mathcal{E}$) and shape prior (defined on the clique set $\mathcal{F}$). The definitions of the data energy $E_{\text{data}}(\mathbf{X})$ and the prior energy $E_{\text{prior}}(\mathbf{X})$ are given as follows.

3.1 Regional & boundary support

The data energy $E_{\text{data}}(\mathbf{X})$ attracts the model to the desired object boundary in terms of image visual properties. We employ region-based and boundary-based measurements to compute the data energy $E_{\text{data}}(\mathbf{X}) = E_{\text{Rg}}(\mathbf{X}) + E_{\text{Bd}}(\mathbf{X})$. The two typical image supports are based on the hypothesis that the object of interest can be distinguished from the background by their statistical properties in an observed image $\mathbf{I}$. Given a model instance $\mathbf{X}$, the image domain Ω is partitioned into the object region $\Omega_{\text{obj}}(\mathbf{X})$ and the background region $\Omega_{\text{bck}}(\mathbf{X})$ according to the model boundary $B(\mathbf{X})$. Let $p_{\text{obj}}, p_{\text{bck}}$ denote the appearance distribution models of the object and the background respectively.

Region-based energy captures the homogeneity properties of different populations observed in the image. Assuming that there is no correlation between the regions labeling and the pixels within each region are independent, it can be computed as follows.

$$\begin{aligned} E_{\text{Rg}}(\mathbf{X}) &= \sum_{i\in\Omega_{\text{obj}}(\mathbf{X})} -\log p_{\text{obj}}(\mathbf{I}_i) + \sum_{i\in\Omega_{\text{bck}}(\mathbf{X})} -\log p_{\text{bck}}(\mathbf{I}_i) \\ &= \sum_{i\in\Omega_{\text{obj}}(\mathbf{X})} -\log \frac{p_{\text{obj}}(\mathbf{I}_i)}{p_{\text{bck}}(\mathbf{I}_i)} + \sum_{i\in\Omega} -\log p_{\text{bck}}(\mathbf{I}_i) \\ &= \sum_{i\in\Omega_{\text{obj}}(\mathbf{X})} -\log \frac{p_{\text{obj}}(\mathbf{I}_i)}{p_{\text{bck}}(\mathbf{I}_i)} + \text{constant} \end{aligned} \tag{10}$$

where $\mathbf{I}_i$ is the image representation (*e.g.* intensity, RGB values or a feature vector) of the pixel/voxel i. Since the integration of the likelihood of the background over the

entire image domain Ω is constant, it can be ignored in the regional energy. In this context, the regional energy is simplified as the integration over the object region, while we denote the integral function $f(\cdot) = -\log \frac{p_{\text{obj}}(\mathbf{I}(\cdot))}{p_{\text{bck}}(\mathbf{I}(\cdot))}$.

Based on the Divergence Theorem, the regional energy can be exactly factorized into higher order terms in MRFs [10]. In 2D cases, the Divergence Theorem states the equivalence between a line integral along a closed curve and a double integral over its bounded region.

$$\iint_{\Omega_{\text{obj}}(\mathbf{X})} f(x,y)dxdy = \oint_{B(\mathbf{X})} F(x,y)dy \tag{11}$$

where the function $f(x,y)$ is the derivative of the function $F(x,y)$ with respect to x. Thus we can compute $F(x,y) = \int_0^x f(t,y)dt$, considering t as the variable. In other words, if we consider an image of the likelihood function $f(x,y)$ over the image domain, then the function $F(x,y)$ over the image domain generates the related integral image with respect to the x axis.

Furthermore, since the model boundary $B(\mathbf{X})$ is composed by a set of line segments determined by two end points, the regional energy of the line integral around the closed curve can be factorized into pair-wise terms:

$$E_{\text{Rg}}^{(1)}(\mathbf{X}) = \sum_{c\in\mathcal{E}} \int_{\mathbf{x}_c} F(x,y)dy \tag{12}$$

Now we deal with the 3D cases, where the Divergence Theorem states that the outward flux of a vector field through a closed surface is equal to the volume integral of the divergence over the region inside the surface.

$$\iiint_{\Omega_{\text{obj}}(\mathbf{X})} f(x,y,z)dxdydz = \oiint_{B(\mathbf{X})} \mathbf{F}\cdot\mathbf{n}\, ds \tag{13}$$

where $f(x,y,z) = \nabla\cdot\mathbf{F}$ is the divergence of the differentiable vector filed $\mathbf{F} = (F_x, F_y, F_z)$, $\mathbf{n}$ is the outward pointing unit normal field of the boundary surface $B(\mathbf{X})$. In our case, the scale-valued function $f(x,y,z)$ is the image likelihood. Let $F_x = F_y = 0$, we have the relation between $F_z(x,y,z)$ and $f(x,y,z)$, where t is the variable.

$$F_z(x,y,z) = \int_0^z f(x,y,t)dt = \int_0^z -\log\frac{p_{\text{obj}}(\mathbf{I}(x,y,t))}{p_{\text{bck}}(\mathbf{I}(x,y,t))}dt \tag{14}$$

Since the boundary surface $B(\mathbf{X})$ is a triangulated mesh in our case, the surface integral over the closed surface of the volume can be factorized into the integral over each triangle region of the mesh:

$$E_{\text{Rg}}^{(2)}(\mathbf{X}) = \sum_{c\in\mathcal{E}} \iint_{\mathbf{x}_c} \mathbf{F}\cdot\mathbf{n}\, ds \tag{15}$$

where the clique set $\mathcal{E}$ consists of triplets of points which compose the triangulated mesh. Given a triplet $(i,j,k)\in\mathcal{E}$, the outward pointing unit normal $\mathbf{n}$ can be computed by the cross product of two vectors $\mathbf{n} = \mathbf{x}_i\mathbf{x}_j \times \mathbf{x}_j\mathbf{x}_k$.

On the other hand, boundary-based support characterizes the discontinuity properties between different regions. It encourages the model boundary to be located on the real boundary between the object and the background in the image. We define the boundary energy as the integral of the appearance discontinuities along the model boundary $B(\mathbf{X})$ which can be decomposed into pairwise or second-order terms.

$$\begin{aligned} E_{\text{Bd}}^{(1)}(\mathbf{X}) &= \sum_{c\in\mathcal{E}} \int_{\mathbf{x}_c} G(x,y)ds \\ E_{\text{Bd}}^{(2)}(\mathbf{X}) &= \sum_{c\in\mathcal{E}} \iint_{\mathbf{x}_c} G(x,y,z)ds \end{aligned} \tag{16}$$

where $E_{\text{Bd}}^{(1)}$ and $E_{\text{Bd}}^{(2)}$ denote the boundary energy respectively for the 2D and 3D cases. The discontinuity function G can be considered as a distance map to the edges. It is acquired by two steps: (1) We apply an edge detector (such as Canny operator) on the observed image to detect the edges, (2) We use distance transform of the edge response to generate the distance map. The map labels each pixel/voxel of the image with the distance to the nearest edge, thus if the pixel/voxel is close to the edges, the function G returns a small value. To minimize the boundary-based energy means that the model boundary is attracted by strong edges that corresponding to locations with local-maxima image gradient values.

3.2 Prior knowledge constraints

The shape prior energy $E_{\text{prior}}(\mathbf{X})$ imposes the geometric constraints of the model in order to produce a valid shape. Based on our sparse graphic shape prior which is modeled by local interactions, the prior energy can be encoded using higher order potentials.

$$E_{\text{prior}}(\mathbf{X}) = \sum_{c\in\mathcal{F}} -w_c \cdot \log p_c(\alpha_c, \beta_c) \tag{17}$$

where $\mathcal{F}$ consists of a set of triplet cliques. Each clique $c \in \mathcal{F}$ is associated with a weight w_c and the probability density p_c of two inner angles from learning. To this end, the total MRF energy can be integrated with the date energy and the prior energy:

$$E(\mathbf{X}) = \sum_{c\in\mathcal{E}} \psi(\mathbf{x}_c) + \sum_{c\in\mathcal{F}} \phi(\mathbf{x}_c) \tag{18}$$

where ψ and ϕ encode respectively the data potential and the prior potential:

$$\begin{cases} \psi^{(1)}(\mathbf{x}_c) = \lambda_1 \cdot \int_{\mathbf{x}_c} F dy + \lambda_2 \cdot \int_{\mathbf{x}_c} G ds \\ \psi^{(2)}(\mathbf{x}_c) = \lambda_1 \cdot \iint_{\mathbf{x}_c} (\mathrm{F}\cdot\mathrm{n}) ds + \lambda_2 \cdot \iint_{\mathbf{x}_c} G ds \\ \phi(\mathbf{x}_c) = -w_c \cdot \log p_c(\alpha_c, \beta_c) \end{cases} \tag{19}$$

We denote $\psi^{(1)}$ and $\psi^{(2)}$ as the data potentials respectively for 2D and 3D cases, while $\lambda_1 > 0$ and $\lambda_2 > 0$ being two weight coefficients. After all the energy terms are defined, we adopt a dual-decomposition optimization framework [12] to perform the Maximum-a-Posteriori (MAP) inference for the proposed higher-order MRF.

4 Experimental Validation

We validate the proposed method in both 2D hand segmentation and 3D left ventricle segmentation. Manual segmentations on the database are available and are considered as ground truth for both learning and validation purposes. An iterative scheme is employed to search for the optimal model instance in the test image. Given an initialized model, the label space of each node is composed by a set of displacements of the current position. The model is updated by the optimal displacements in each iteration, while the displacement set is adapted to a coarse to fine setting during the model deformation. The experiments were run on a 2.8GHz, Quad Core, 12GB RAM computer.

4.1 2D hand segmentation

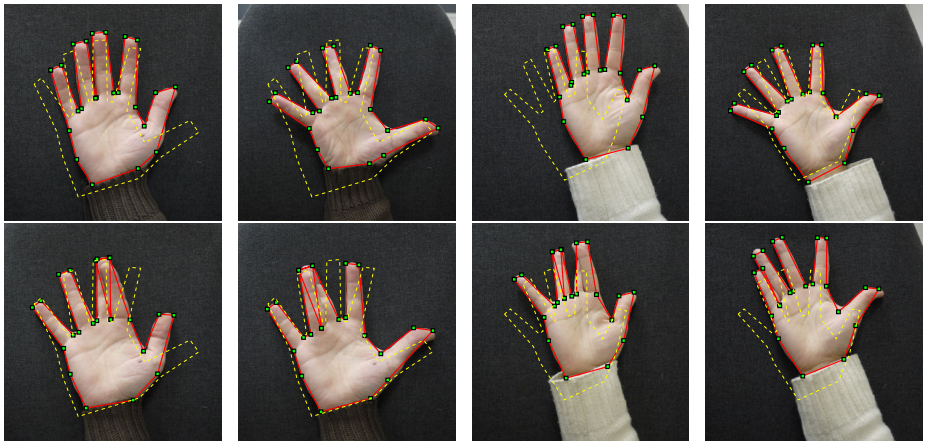
Our 2D hand dataset consists of 40 right hand examples with different poses and movements between the fingers. The shape model consists of 23 control points, and the MRF learning is performed on the complete graph of 1771 triplet cliques. The learning method can efficiently deal with higher-order MRFs, as shown in Fig.2 (a) where the objective function (5) converges in less than one minute in the learning procedure. Given the sparsity property of the learned parameter vector $\mathbf{w}$ as shown in Fig.2 (b), we chose a number of 100 cliques with the largest parameters to represent the shape prior.

Some segmentation results of our knowledge-based method are shown in Fig.3, where the red solid contours represent our result and the yellow dashed contours represent the initializations. As can be seen, our results are robust to the noise, partial occlusions and complicated background. For example, in the second row where the fingers are partially self-occluded, our method shows the ability to deal with the shapes which have not been seen during training. In the third row, the same images from the first row are artificially added with Gaussian noise and black obstructions, while we deal with these cases with a larger weight of the prior energy, which is also the reason why a part of sleeve is mis-labeled as the hand in the second image. The fourth row shows our results on a set of video images with complicated background.

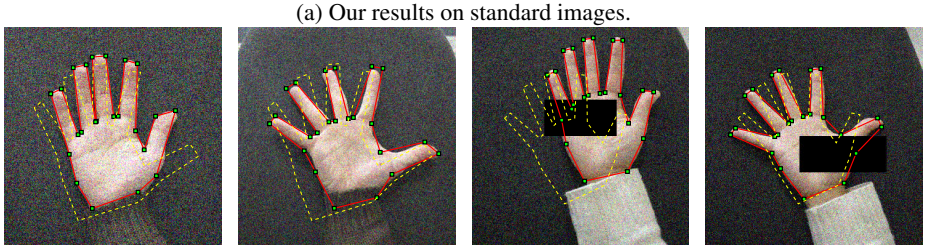
For both quantitative and comparison purposes, we compare our method with Active Shape Model (ASM) in Fig.4 (a). The dice coefficients, the similarity measurements of the result and the ground truth, verify our better performance than ASM. Moreover, benefit from the sparse graphic shape prior, our segmentation takes 20 seconds per image while the one using complete graph takes more than 4 minutes.

4.2 3D left ventricle segmentation

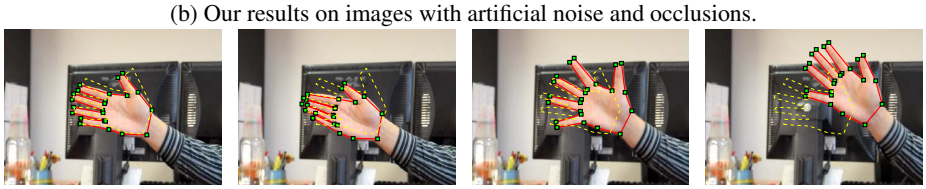
A dataset of 20 3D CT cardiac images is used to validate the proposed method in 3D segmentation application. The shape model consists of 88 control points on the myocardium surface as well as the atrium surface (see Fig.1 (a)), and 172 triangle faces producing the surface mesh (see Fig.1 (b)). The sparse graphic shape priors are composed by 1000 triplet cliques selected from the MRF learning. Regarding the image support, each voxel is represented by a feature vector consisting of patches of intensities centered at the voxel and Gabor features. The appearance models of the object and the background are learned by Adaboost classifiers.



(a) Our results on standard images.



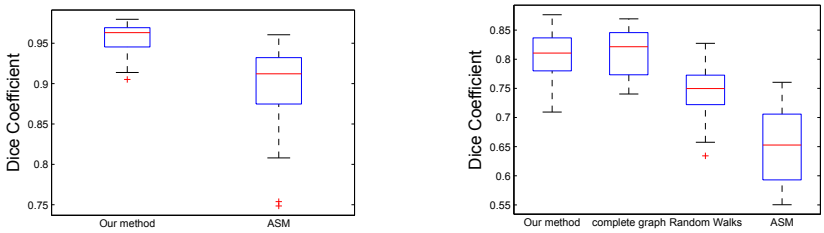
(b) Our results on images with artificial noise and occlusions.



(c) Our results on video images with cluttered background.

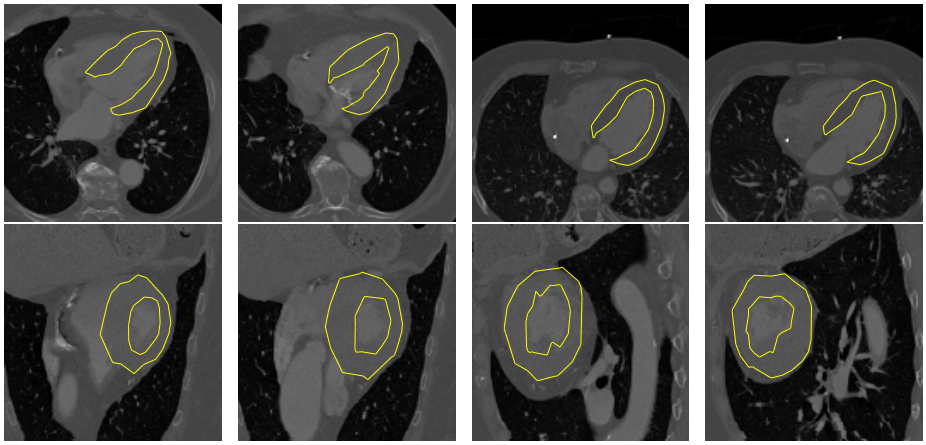
Fig. 3. 2D hand segmentation results.

We perform a leave-one-out cross-validation on the dataset. Some segmentation results are shown in the Fig.5, where the yellow contours represent our results, and the green contours represent the ASM results. As can be observed, our results exhibit better accuracy on the boundary as well as being robust to the papillary muscles in the blood pool, while ASM is easily trapped in a local minimum. In addition, we compare our method with other methods for quantitative evaluation in the Fig.4 (b). From left to right, we present the Dice coefficients obtained by our sparse graph model, the complete graph model [10], the Random Walks algorithm [7] and the standard ASM [2]. The method [10] also shows good performance, but it suffers from high computational complexity introduced by the complete graph (hours per volume segmentation). Thus it is limited when an increasing number of control points is required. Our method is more efficient with reduced computation complexity in both energy computation and optimization process (15 minutes per volume segmentation).

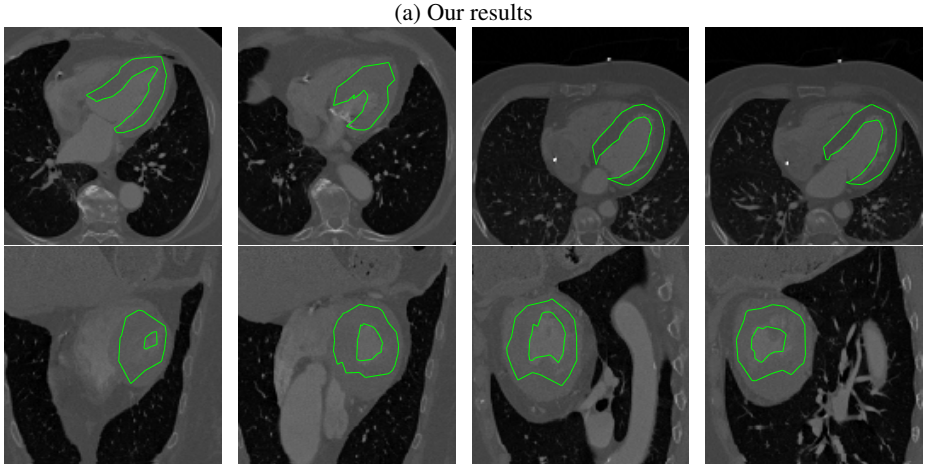


(a) 2D hand segmentation (b) 3D heart segmentation

Fig. 4. Quantitative results of the dice coefficients.



(a) Our results



(b) ASM results

Fig. 5. 3D segmentation results on cardiac CT volumes.

5 Conclusion

In this paper we have studied the problem of knowledge-based object segmentation. The main contribution of our method is a novel L_1 sparse higher order graph representation that is pose invariant towards modeling shape variations. This formulation consists of a compact representation that captures shape variations through accumulation of local constraints. Furthermore, it eliminates the need of bringing shapes to the same reference space, either towards building an appropriate statistical model or during inference towards creating consistency between the space on where statistics were learned and the image to be segmented. This model has been endowed with conventional graph connectivity allowing the natural use of boundary and regional image support.

Pixel/voxel-based are well studied/popular approaches withing graph-theoretic methods for segmentation. Coupling them with higher-order priors will have a double benefit. First, pixel-based methods will become robust with respect to noise and occlusions while prior-based methods (as the one presented here) would require a less dense point distribution model, since capturing fine shape variations could be achieved through the pixel/voxel-based approach.

This is the main future direction of our work through a unified formulation that is solved using an one-shot optimization framework. The application of these methods to cardiac segmentation over the entire cardiac cycle on magnetic resonance and computed tomography images is currently under investigation.

References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. IJCV **1** (1988) 321–331
2. Cootes, T., Taylor, C., Cooper, D., Graham, J., et al.: Active shape models-their training and application. Computer Vision and Image Understanding **61** (1995) 38–59
3. Chen, Y., Tagare, H., Thiruvenkadam, S., Huang, F., Wilson, D., Gopinath, K., Briggs, R., Geiser, E.: Using prior shapes in geometric active contours in a variational framework. IJCV (2002)
4. Rousson, M., Paragios, N.: Prior knowledge, level set representations & visual grouping. IJCV (2008)
5. Cremers, D., Osher, S., Soatto, S.: Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. IJCV (2006)
6. Freedman, D., Zhang, T.: Interactive graph cut based segmentation with shape priors. In: CVPR. Volume 1. (2005) 755–762
7. Grady, L.: Random Walks for Image Segmentation. PAMI **28** (2006) 1768–1783
8. Seghers, D., Hermans, J., Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P.: Model-based segmentation using graph representations. In: MICCAI. (2008)
9. Besbes, A., Komodakis, N., Langs, G., Paragios, N.: Shape priors and discrete mrfs for knowledge-based segmentation. In: CVPR. (2009)
10. Xiang, B., Wang, C., Deux, J., Rahmouni, A., Paragios, N.: 3d cardiac segmentation with pose-invariant higher-order mrfs. In: ISBI. (2012) 1425–1428
11. Komodakis, N.: Efficient training for pairwise or higher order crfs via dual decomposition. In: CVPR. (2011) 1841–1848
12. Komodakis, N., Paragios, N., Tziritas, G.: MRF optimization via dual decomposition: Message-passing revisited. In: ICCV. (2007)