3D Knowledge-based Segmentation Using Pose-Invariant Higher-Order Graphs

Chaohui Wang[1,2], Olivier Teboul[1,3], Fabrice Michel[1,2], Salma Essafi[1,2], and Nikos Paragios[1,2]

[1] Laboratoire MAS, Ecole Centrale de Paris, France
[2] Equipe GALEN, INRIA Saclay - Île de France, Orsay, France
[3] Microsoft France
chaohui.wang@ecp.fr

Abstract. Segmentation is a fundamental problem in medical image analysis. The use of prior knowledge is often considered to address the ill-posedness of the process. Such a process consists in bringing all training examples in the same reference pose, and then building statistics. During inference, pose parameters are usually estimated first, and then one seeks a compromise between data-attraction and model-fitness with the prior model. In this paper, we propose a novel higher-order Markov Random Field (MRF) model to encode pose-invariant priors and perform 3D segmentation of challenging data. The approach encodes data support in the singleton terms that are obtained using machine learning, and prior constraints in the higher-order terms. A dual-decomposition-based inference method is used to recover the optimal solution. Promising results on challenging data involving segmentation of tissue classes of the human skeletal muscle demonstrate the potentials of the method.

1 Introduction

Knowledge-based segmentation consists in recovering a region of anatomical interest in a new image. The process often combines data support with manifold learning on the space of adequate solutions. The data term is usually either edge-based or region driven. In the first case, one seeks to position the solution onto pixels exhibiting important intensity discontinuities, which is achieved through a weighted surface integral. Region-based methods assume that the object and the background have distinct statistical properties and seek to create a partition that maximizes the posterior probability density with respect to them. Both methods fail to address the case of anatomical regions of interest being part of the same tissue class, where either edges are not present or statistical separation is not feasible. The case of calf muscle segmentation is a typical example.

Prior knowledge is often encoded through constraining the solution space. This is achieved either through linear models describing the variation of the training set (ASMs) [1], or through projection and minimization of the distance to the learned manifold. The use of simple statistical means (atlas-based methods) [2], parametric or non-parametric priors considered in an explicit [3] or

implicit level set formulation [4] were considered. Such techniques exhibit two important limitations, the first is related to pose invariance and the second is related to their ability to capture statistics on high-dimensional spaces from a small number of training examples. The pose estimation problem arises both in the training and in the inference steps and introduces certain bias on the segmentation process since data are often to be registered in the reference space. The ratio samples versus dimensionality of representations is also a well-known problem in medical imaging.

In this paper, we propose a novel approach that is able to address segmentation for challenging data sets while being pose invariant and being able to capture local variations from small training sets. To this end, we employ a higher-order Markov Random Field (MRF) formulation. The representation of the shape is a point distribution model (PDM) that is used to provide the entire surface through conventional interpolation algorithms. Prior knowledge is modeled through the use of higher-order relative statistics of the PDM. These are invariant to rotation and scale while they can be learned from a small number of training examples. The entire manifold is described through the accumulation of such local constraints. This representation is associated with the randomized forest [5] learning approach that provides an efficient detection algorithm for points of interest exhibiting some statistical properties. These properties can be derived from Gabor-filter-based scale/rotation invariant features. In order to optimize the higher-order MRF's energy, we propose to decompose the original problem, which is difficult to solve directly, into a series of sub-problems each of which corresponds to a factor tree [6,7]. The inference in a factor tree can be done exactly and very efficiently using max-product belief propagation algorithm [7]. In order to evaluate the performance of the method, we have considered the challenging application of segmentation of the calf muscle. The closest work refers to the segmentation algorithms in [8,9,10]. Opposite to these algorithms, our approach explores the 3D space and higher-order interactions between the model variables, inherits natural invariance with respect to the global pose (opposite to [9]), exploits invariant features with respect to this pose (opposite to [9,10]), builds pose invariant statistics (opposite to [8,9,10]) and provides an one-shot optimization result that does not depend on the initial conditions or the reference pose parameters.

The reminder of this paper is organized as follows: we present the higher-order shape prior, the MRF segmentation formulation and the MRF optimization method in section 2. The experimental validation composes section 3, while section 4 concludes the paper.

2   Knowledge-based Segmentation

2.1   Shape Modeling

The shape model consists of a set $\mathcal{V}$ of control points/landmarks that are located on the boundary of the object (Fig. 1(a)). Let x_i ($i \in \mathcal{V}$), a 3-dim vector, denote the 3d position of landmark i and $\mathbf{x} = (x_i)_{i \in \mathcal{V}}$ denote the position of

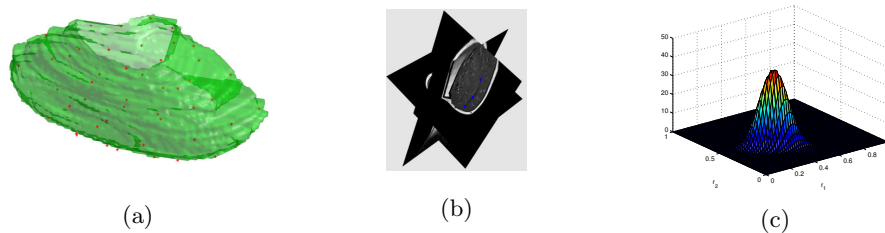(a)   (b)   (c)

Fig. 1. Shape Model. (a) Distribution of the landmarks on the muscle boundary. (b) Two perpendicular slices with a triplet of landmarks (the blue asterisks). (c) The learned Gaussian distribution on $\mathbf{r}_c$ for the triplet shown in (b).

all the landmarks. Given a training set $\mathcal{T}$ composed of M aligned shapes, i.e., $\mathcal{T} = \{\mathbf{x}^{(m)}\}_{m \in \{1,2,...,M\}}$ [11], we aim to learn a prior probability distribution on the different configuration of the shape model from the training set. Instead of learning the statistics on the pair of landmarks (e.g., [8]), which is not appropriate to get a scale-invariant modeling, we propose to learn the statistics of the measurements that are scale-invariant. To this end, let us consider a triplet of points, $c \in \mathcal{C} = \{(i,j,k)|i,j,k \in \mathcal{V} \text{ and } i \neq j, j \neq k, k \neq i\}$ and learn statistics on the relative lengths $(r_{i_c j_c}, r_{j_c k_c}, r_{k_c i_c})$ of the sides [12], which are defined as (take $r_{i_c j_c}$ for example): $r_{i_c j_c} = l_{i_c j_c}/(l_{i_c j_c} + l_{j_c k_c} + l_{k_c i_c})$, where $l_{i_c j_c}$ denotes the Euclidean distance between points i_c and j_c. An important advantage to use the relative lengths is that they can be computed much faster than the angle measurements which are also scale-invariant. For a triplet, it is sufficient to only consider the relative lengths of two sides since the third one is a linear combination of them (i.e., $r_{k_c i_c} = 1 - r_{i_c j_c} - r_{j_c k_c}$). Thus, without loss of generality, we use a Multivate Gaussian Distribution (Fig. 1(c)) to model the distribution of $\mathbf{r}_c = (r_{i_c j_c}, r_{j_c k_c})^T$, i.e., $p_c(\mathbf{r}_c) = \mathcal{N}(\mathbf{r}_c|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$, where the mean $\boldsymbol{\mu}_c$ and the covariance matrix $\boldsymbol{\Sigma}_c$ can be learned from the training set by maximum likelihood:

$$\boldsymbol{\mu}_c = \frac{1}{M} \sum_{m=1}^{M} \mathbf{r}_c^{(m)} \tag{1}$$

$$\boldsymbol{\Sigma}_c = \frac{1}{M} \sum_{m=1}^{M} (\mathbf{r}_c^{(m)} - \boldsymbol{\mu}_c)(\mathbf{r}_c^{(m)} - \boldsymbol{\mu}_c)^T \tag{2}$$

Finally, we get the higher-order shape model $\mathcal{P} = (\mathcal{V}, \mathcal{C}, \{\mathcal{N}(\cdot|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)\}_{c \in \mathcal{C}})$, where $\mathcal{V}$ and $\mathcal{C}$ determine the topology of the model while $\{\mathcal{N}(\cdot|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)\}_{c \in \mathcal{C}}$ characterizes the statistic geometry constraints between the triplets.

2.2 Landmark Candidate Detection

Given such a prior model, segmentation can be viewed as finding a mapping between the model points and a new volume. This can be expressed as a correspondence problem that requires detection for the model points in the image or

finding a set of correspondences for each point i ($i \in \mathcal{V}$). To this end, we first learn a classifier for each landmark, and then compute a score for each possible localization, and finally select the L positions that have the best scores. In the experiments, we adopt Randomized Forest [5] to perform the classification.

A randomized forest is composed of a set of T random decision trees. In the decision trees, an internal node consists of a random test on an input feature vector, the result (true or false) of which decides which (left or right) child node the feature vector goes to. And a leaf consists in a histogram $h = (h_1, \dots, h_W)$ (W is the number of classes), which is obtained during the training phase by counting the number of labeled feature vectors that arrive at this leaf. During the testing phase, an unlabeled feature vector is dropped in each decision tree τ and reaches the leaf l_τ, and the normalized histogram of l_τ provides a probability estimation for the feature vector belonging to each class w:

$$P(w|l_\tau) = \frac{h_w}{\sum_i h_i} \tag{3}$$

Finally, we average the probabilities of all the trees to obtain the probability over the forest:

$$P(w|(l_1, \dots, l_T)) = \frac{1}{T} \sum_\tau P(w|l_\tau) \tag{4}$$

We consider all the voxels in a volume as the possible localization of the landmarks. Each voxel is associated with a feature vector which is used as the input for the classifiers. Different features can be considered towards achieving a high quality detection. In order to well capture the local image structure information, we can use a series of 3d Gabor filters [13] with different scale, rotation parameters. We adopt the method proposed in [14] to sample these parameters such that scaling/rotation of the image becomes a translation of these parameters and then estimate the Fourier Transform Modulus (FTM) of the filter output to eliminate variations due to these translations (because the FTM is translation invariant). Due to the symmetry of FTM, it is enough to consider only half of the FTM domain by removing the redundant coefficients. In such a way, we get a feature vector for each voxel that is scale and rotation invariant.

Fig. 2 shows the detected candidate results for four landmarks at different locations on a testing muscle data.

2.3   Higher-order MRF Segmentation Formulation

The shape model, together with the evidence from the image support, is formulated within a higher-order MRF towards image segmentation.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{C})$ denote a hypergraph with a node set $\mathcal{V}$ and a clique set $\mathcal{C}$, $\mathbf{U}^G = \{U_q(\cdot)\}_{q \in \mathcal{V}}$ the singleton potentials defined on the node set $\mathcal{V}$, and $\mathbf{H}^G = \{H_c(\cdot)\}_{c \in \mathcal{C}}$ the clique potentials defined on the clique set $\mathcal{C}$. And then let $\mathrm{MRF}^{\mathcal{G}}(\mathbf{U}^{\mathcal{G}}, \mathbf{H}^{\mathcal{G}})$ denote a higher-order MRF with topology $\mathcal{G}$ as well as the potentials $\mathbf{U}^G$ and $\mathbf{H}^G$. In our problem, we associate a landmark to a node i

($i \in \mathcal{V}$), a triplet to a third-order clique c ($c \in \mathcal{C}$), with one-to-one mappings[4]. The latent variable X_i corresponding to node i is a 3-dim vector denoting the position of node/point i. $\mathbf{x}^{\mathcal{G}}$ represents the configurations of all the nodes, i.e. $\mathbf{x}^{\mathcal{G}} = (x_i)_{i \in \mathcal{V}}$. The candidate set of each variable is denoted by $\mathcal{X}_i$ ($i \in \mathcal{V}$). Thus the candidate set $\mathcal{X}^{\mathcal{G}}$ of all the variable of the MRF is defined as: $\mathcal{X}^{\mathcal{G}} = \prod_{i \in \mathcal{V}} \mathcal{X}_i$, which composes all the possible configurations of the shape model. The candidate set $\mathcal{X}_i$ for each node consists of the detected landmark candidates (section 2.2).

The segmentation problem is transformed into estimating the optimal positions of the landmarks, i.e., the optimal configuration $\mathbf{x}^{\mathcal{G}}_{\mathrm{opt}}$ of MRF$^{\mathcal{G}}$, which is formulated as a minimization of the MRF's energy $E(\mathbf{x}^{\mathcal{G}})$:

$$\mathbf{x}^{\mathcal{G}}_{\mathrm{opt}} = \arg \min_{\mathbf{x}^{\mathcal{G}} \in \mathcal{X}^{\mathcal{G}}} E(\mathbf{x}^{\mathcal{G}}) \tag{5}$$

where the energy of MRF$^{\mathcal{G}}$ is defined as:

$$E(\mathbf{x}^{\mathcal{G}}) = \sum_{q \in \mathcal{V}} U_q(x_q) + \sum_{c \in \mathcal{C}} H_c(\mathbf{x}_c) \tag{6}$$

where $\mathbf{x}_c$ denotes the configuration $(x_q)_{q \in c}$ of clique c. The singleton potentials $\mathbf{U}^G$ and third-order clique potentials $\mathbf{H}^G$ are presented as follows:

The singleton potential $U_q(x_q)$ ($q \in \mathcal{V}$) is the negative log-likelihood which imposes penalty for the landmark q being located at position x_q in image $\mathbf{I}$:

$$U_q(x_q) = - \log p(\mathbf{I}|x_q) \tag{7}$$

where we define $p(\mathbf{I}|x_q)$ using the classifier's output probability value for landmark q being located at x_q (Eq. 4).

The higher-order clique potential $U_c(\mathbf{x}_c)$ ($c \in \mathcal{C}$) encodes the statistic geometry constraints between the triplet c of points and is defined as:

$$U_c(\mathbf{x}_c) = -\alpha \cdot \log p_c(\mathbf{r}_c(\mathbf{x}_c)) \tag{8}$$

where $\alpha > 0$ is a weight coefficient, $\mathbf{r}_c(\mathbf{x}_c)$ denotes the mapping from the position of the triplet c to the 2-dim relative lengths of the sides, and the prior distribution p_c has been defined in section 2.1.

2.4   MRF Decomposition & Optimization

In order to perform inference in the proposed higher-order MRF, we adopt the well-known dual-decomposition optimization framework [15,16]. The key idea of such a framework is: instead of minimizing directly the energy of the original problem, we maximize a lower bound on it by solving the dual to the linear programming (LP) relaxation [15].

To this end, we first decompose the original problem (corresponding to MRF$^{\mathcal{G}}$) into a set of sub-problems (corresponding to $\{\text{MRF}^{\mathcal{G}_s}\}_{s \in \mathcal{S}}$, where $\mathcal{S}$ is the set

---

[4] Due to such one-to-one mappings, in this section, we reuse the notation $\mathcal{V}$ and $\mathcal{C}$ to denote the node set and the clique set, respectively.
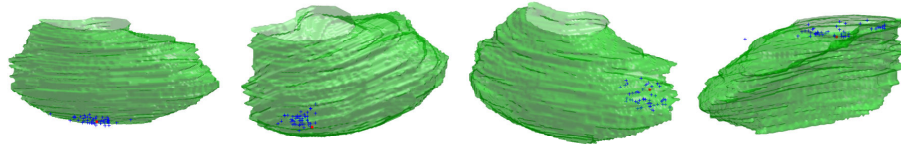
Fig. 2. Landmark detection results. The red hexagram represents the ground truth while the blue plus signs represent 50 detected candidates. The reference segmentation surface is provided to visually measure the distance between candidates and the ground truth.

of sub-problems), each of which is smaller and solvable. More concretely, we decompose the original hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{C})$ into a set of sub-hypergraphs $\{\mathcal{G}_s = (\mathcal{V}_s, \mathcal{C}_s)\}_{s \in \mathcal{S}}$ such that $\mathcal{V} = \cup_{s \in \mathcal{S}} \mathcal{V}_s$ and $\mathcal{C} = \cup_{s \in \mathcal{S}} \mathcal{C}_s$. In order to form $\{\text{MRF}^{\mathcal{G}_s}\}_{s \in \mathcal{S}}$ corresponding to the sub-problems, the original MRF potentials are decomposed into the sub-hypergraphs such that $\mathbf{U} = \sum_{s \in \mathcal{S}} \mathbf{U}^{\mathcal{G}_s}$ and $\mathbf{H} = \sum_{s \in \mathcal{S}} \mathbf{H}^{\mathcal{G}_s}$, which can be achieved simply by setting $U_q^{\mathcal{G}_s} = \frac{U_q^{\mathcal{G}}}{|\{s | q \in \mathcal{V}_s\}|}$ and $H_c^{\mathcal{G}_s} = \frac{H_c^{\mathcal{G}}}{|\{s | c \in \mathcal{C}_s\}|}$. The dual-decomposition [15] states that the sum of the minimum energies of the sub-problems provides a lower bound to the minimum energy of the original MRF. Furthermore, the problem of maximizing such a lower bound over its feasible set is then convex. Like [16], we adopt a projected subgradient method to perform this maximization so as to combine the solution of the sub-problems to get the solution of the original problem.

The most challenging component is how to define the sub-problems to decompose the original problem[5]. For the purpose of solving the inference, we adopt *factor graph* [6,7] to represent the MRFs. To this end, we introduce additional nodes for the factors corresponding to the singleton potentials $U_q(x_q)$ ($q \in \mathcal{V}$) and the third-order clique potentials $H_c(\mathbf{x}_c)$ ($c \in \mathcal{C}$). Considering both the complexity and the quality of the optimum, we propose to decompose the original factor graph into a series of factor trees (i.e., factor graphs without loop) such that a higher-order clique factor appears in one and only one factor tree. The inference in a factor tree can be done exactly and very efficiently using max-product belief propagation algorithm [7] with complexity $O(NL^K)$, where N, L and K denote the number of nodes, the number of candidates for each node, and the maximum order ($K = 3$) of the factors, respectively.

3 Experimental Results

We used the data set that was previously used in [17] to validate the proposed method. This data set consists of 25 3D MRI subjects whose calf part was imaged. The voxel spacing is of $0.7812 \times 0.7812 \times 4$ mm and each volume consists

---

[5] We cannot use the decomposition scheme proposed in [16], since the higher-order clique potentials are not pattern-based.

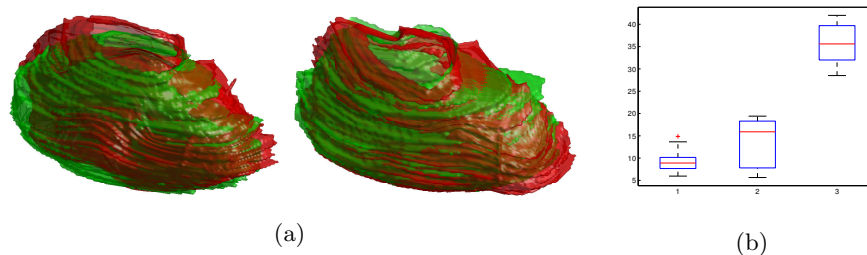(a)   (b)

Fig. 3. Experimental Results. (a) Surface reconstruction results (green: reference segmentation. red: reconstruction result). (b) Boxplots of the average landmark error measure in voxel (1. our method. 2. method in [17]. 3. standard ASM method.). On each box, the central mark in red is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points.

of 90 slices of 4mm thickness acquired with a 1.5 T Siemens scanner. Standard of reference was available, consisting of annotations provided by experts for the Medial Gastrocnemius (MG) muscle.

We performed a leave-one-out cross validation on the whole data set. For comparison purpose, we considered as alternative segmentation methods[6] the ones presented in [17]. We present in Fig. 3(a) the surface reconstruction results using the estimated position of the landmarks and thin plate spline (TPS), while in Fig. 3(b) the average distance between the real landmark position and the one estimated from our algorithm, and the ones reported in [17] including the one obtained using standard active shape models. We reduce landmark localization error by an average of factor 2 in terms of voxel error compared to [17] that is considered to be the state of the art. The analysis of the results shows that the proposed prior and the inference using higher-order graphs globally perform well while the main limitation is introduced from the landmark candidate detection process. Since the method establishes correspondences between the model and the detected landmarks, in the absence of meaningful candidates the method fails to recover optimally the global shape. Regarding computational complexity, the method is linear with respect to the number of higher-order cliques and cubic with respect to the number of landmarks candidates per point.

4   Conclusion

In this paper, we have proposed a novel approach for 3D segmentation using pose invariant higher-order MRFs. Our method models the prior manifold through accumulation of local densities involving pose invariant combinations of points. Segmentation is expressed as a higher-order MRF optimization problem, where

---

[6] Opposite to [17], we have considered a subset of 50 from the 895 model landmarks uniformly distributed in the model-space (Fig. 1).

machine learning techniques and pose invariant features are considered to determine candidate positions for the model points. Promising results that clearly outperform the prior art in very challenging data sets demonstrate the potentials of the method.

Extending the framework to deal with missing correspondences is the most promising direction to overcome the challenge of correctly estimating the position of all the landmarks of the model. Redundancy is a natural property inherited from the exhaustive construction of the higher-order model. The optimization of the graph connectivity towards reducing the computational complexity of the method is a straightforward direction as suggested in [8] through dimensionality reduction on the graph space. Last, but not least the case of spatio-temporal higher-order priors on anatomical structures with dynamic behavior is currently under investigation.

References

1. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. Comput. Vis. Image Underst. **61**(1) (1995) 38–59
2. Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J.: A probabilistic atlas of the human brain: Theory and rationale for its development : The international consortium for brain mapping (icbm). NeuroImage **2**(2, Part 1) (1995) 89 – 101
3. Staib, L.H., Duncan, J.S.: Boundary finding with parametrically deformable models. IEEE Trans. Pattern Anal. Mach. Intell. **14**(11) (1992) 1061–1075
4. Leventon, M.E., Grimson, W.E.L., Faugeras, O.D.: Statistical shape influence in geodesic active contours. In: CVPR. (2000)
5. Breiman, L.: Random forests. Machine Learning **45** (2001) 5–32
6. Frey, B.J.: Graphical Models for Machine Learning and Digital Communication. MIT Press (1998)
7. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (August 2006)
8. Besbes, A., Komodakis, N., Langs, G., Paragios, N.: Shape priors and discrete mrfs for knowledge-based segmentation. In: CVPR. (2009)
9. Seghers, D., Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P.: Minimal shape and intensity cost path segmentation. IEEE TMI **26**(8) (2007) 1115–1129
10. Donner, R., Langs, G., Bischof, H.: Sparse mrf appearance models for fast anatomical structure localisation. In: British Machine Vision Conference. (2007)
11. Luo, B., Hancock, E.R.: Iterative procrustes alignment with the em algorithm. Image Vision Comput. **20**(5-6) (2002) 377–396
12. Lekadir, K., Merrifield, R., Yang, G.: Outlier detection and handling for robust 3-d active shape models search. IEEE TMI **26** (2007) 212–222
13. Bernardino, A., Santos-Victor, J.: Fast iir isotropic 2-d complex gabor filters with boundary initialization. Volume 15. (2006) 3338–3348
14. Kokkinos, I., Yuille, A.: Scale invariance without scale selection. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). (2008)
15. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific (1999)
16. Komodakis, N., Paragios, N.: Beyond pairwise energies: Efficient optimization for higher-order mrfs. In: CVPR. (2009)
17. Essafi, S., Langs, G., Deux, J.F., Rahmouni, A., Bassez, G., Paragios, N.: Wavelet-driven knowledge-based mri calf muscle segmentation. In: Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging. (2009)