# Joint Model-Pixel Segmentation with Pose-Invariant Deformable Graph-Priors

Bo Xiang[1,2], Jean-Francois Deux[3], Alain Rahmouni[3], and Nikos Paragios[1,2]

[1] Center for Visual Computing, Ecole Centrale de Paris, France
[2] Equipe GALEN, INRIA Saclay, Île-de-France, France
[3] Radiology Department, Henri Mondor Hospital, Créteil, France

**Abstract.** This paper proposes a novel framework for image segmentation through a unified model-based and pixel-driven integrated graphical model. Prior knowledge is expressed through the deformation of a discrete model that consists of decomposing the shape of interest into a set of higher order cliques (triplets). Such decomposition allows the introduction of region-driven image statistics as well as pose-invariant (*i.e.* translation, rotation and scale) constraints whose accumulation introduces global deformation constraints on the model. Regional triangles are associated with pixels labeling which aims to create consistency between the model and the image space. The proposed formulation is pose-invariant, can integrate regional statistics in a natural and efficient manner while being able to produce solutions unobserved during training. The challenging problem of tagged cardiac MR image segmentation is used to demonstrate the performance potentials of the method.

## 1 Introduction

Segmentation is one of the most well studied topics in computer vision. Model-free methods are often based on clustering, aiming at grouping together pixels with consistent intensity properties. Knowledge-driven methods, on the other hand, aim to find a solution that is a compromise between the one produced from the observations and the one expressed from the model space.

Popular examples of model-free segmentation refer to the mean-shift method [4], variational formulations such as the Mumford-Shah framework [12], or graph-based methods including normalized cuts [9], graph-cuts [2]. Due to the lack of assumptions on the geometric form of the object of interest, these methods are rather flexible in terms of spread of admissible solutions while it can also lead to erroneous results due to intensity variability, occlusions, noise presence.

Knowledge-based methods are either manifold constrained or manifold enhanced. The former class of methods models geometric variation of the object of interest and then seeks an instance of this space in the image. Active shape/appearance models [6,5] and atlas-based methods [7] are popular examples. Manifold enhanced methods aim to minimize the distance of the solution from the learned manifold, *e.g.* active contours/surface models [10,11]. Both classes of methods inherit a severe limitation with respect to pose, due to the

K. Mori et al. (Eds.): MICCAI 2013, Part III, LNCS 8151, pp. 267–274, 2013.

fact that the current solution should be brought to the same referential as the ones used in learning.

Recently, [14] proposed a pose invariance model through a higher order graph-based formulation with promising results. Due to the discrete shape model, it might produce significant segmentation errors on the boundary. Meanwhile, combined image-model segmentation has been investigated as well. The approach of [13] considered extremely simple shape priors. [3] addressed the problem within an alternating minimization approach involving both discrete and continuous optimization process, where no guarantees on the optimality properties of the obtained solution could be satisfied.

In this paper, we propose a novel pose-invariant segmentation approach that simultaneously solves the problem in both model space and image space. It is achieved by the definition of an objective function aiming to: (i) assign labels to image pixels in order to maximize the image likelihood [2], (ii) deform a point distribution model in order to maximize the geometric likelihood of the model as well as the model-to-image likelihood [14], (iii) impose consistency between the two label spaces. The resulting higher order graphical model formulation is solved by using a state of the art message passing algorithm [8]. Promising results on a challenging clinical setting demonstrate the potentials of our method.

The remainder of the paper proceeds as follows. We present the shape model in section 2 and the segmentation energy is defined in section 3. Experimental validation are shown in section 4 while section 5 concludes the paper.

## 2 Shape Representation

We adopt the pose invariant shape model in [14]. It consists of: (1) a set of control points distributed on the boundary (Fig.1 left); (2) a set $\mathcal{C}$ of cliques including all possible combinations of three points; (3) a series of probability distributions $p_c$ of each triplet $c \in \mathcal{C}$ learned from a training set.

Let $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{n}$ denote a shape instance defined by $n$ points, where $\mathbf{x}_i$ denotes the coordinates of point $i$. Given a training set of $K$ samples $\{\mathbf{X}^t\}_{t=1}^{K}$, we assume that point correspondences exist within the training set, but no need to align all the samples in a common reference space. For a triplet clique $c = (i, j, k) \in \mathcal{C}$, the co-occurrence probability of the three points $\mathbf{x}_c = (\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$ can be represented by their inner angles $\{\alpha_c = \angle \mathbf{x}_j\mathbf{x}_i\mathbf{x}_k, \beta_c = \angle \mathbf{x}_i\mathbf{x}_j\mathbf{x}_k\}$.

$$p_c(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) = p_c(\alpha_c, \beta_c) \tag{1}$$

This angle representation of triplet is invariant to global pose (*i.e.* translation, rotation, scale) of the shape of the object. With $K$ training samples, the probability density distributions $p_c(\alpha_c, \beta_c)$ of triplet $c$ are learned using a standard probabilistic model (*e.g.* Gaussian distribution model). Assuming that the local constraints on triplets are independent, the global shape is constrained through the accumulation of all the local constraints.

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

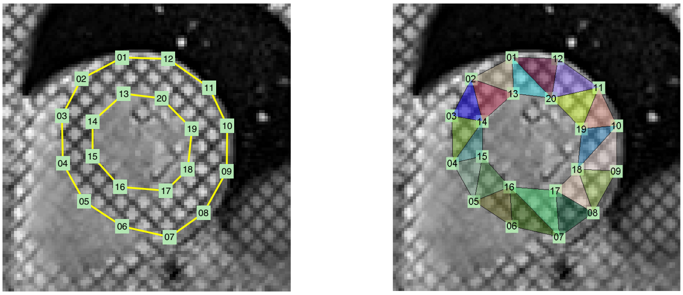


**Fig. 1.** Shape model. Left: point distribution model. Right: model triangulation.

The shape model inherits pose invariance from the local representation. As a result, no shape alignments to the same referential are needed for both training samples and testing shapes. Moreover, it can be easily encoded with a MRF inference due to the local interactions defined by *prior related* clique set $\mathcal{C}$.

In addition to prior related concerns, we introduce a *data related* clique set $\mathcal{A}$ which decomposes the object region into triangles (Fig.1 right). Such a triangulation can be applied to any shape (heart, liver etc.) represented as a polygonal area and it should meet the conditions: (1) each triplet is a subset of the model and its corresponding triangle region should be included only in the object; (2) these triangle regions should not overlap; (3) the union of these triangle regions recovers the whole object domain. Using model triangulation facilitates to factorize the regional-driven energy as well as to introduce pixel and model interactions which will be shown in the following segmentation framework.

## 3 Combined Model-Pixel Based Segmentation

In this section, we propose a framework to combine both model-based and pixel-based segmentation. The aim is to simultaneously deform the shape model to an observing image and to label the image pixels as object/background within an interconnected graphical model.

### 3.1 MRF Formulation

Now we address the segmentation problem within a higher order Markov Random Field (MRF) formulation. The proposed graph model $G$ consists of two sub-graphs: (1) The model-based $G_m$ consists of a set $\mathcal{V}_m = \{1, \cdots, n\}$ of *model* nodes (associated with $n$ points in shape model) and a set of cliques $\mathcal{C}_m$ used in model-based segmentation independently; (2) The pixel-based $G_p$ consists of a set $\mathcal{V}_p = \{1, \cdots, k\}$ of *pixel* nodes (associated with $k$ pixels in the image) and a set of cliques $\mathcal{C}_p$ introduced by pixel-based segmentation. Moreover, the two sub-graphs are connected with a clique set $\mathcal{C}_{int}$. We illustrate the graph structure in Fig.2, where the yellow upper part represents model-based $G_m$ and the green lower part represents pixel-based $G_p$.

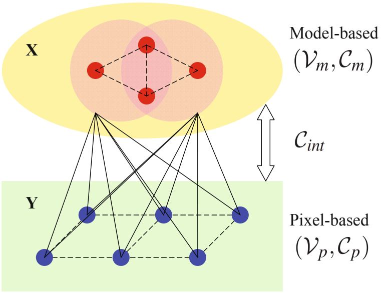

X
Model-based
$(\mathcal{V}_m, \mathcal{C}_m)$
$\mathcal{C}_{int}$
Y
Pixel-based
$(\mathcal{V}_p, \mathcal{C}_p)$

**Fig. 2.** MRF graphical model coupling the model space and the label space

Let $X_{i\in\mathcal{V}_m}$ (*i.e.* point coordinates) and $Y_{i\in\mathcal{V}_p}$ (*i.e.* pixel label) denote the latent random variables for model nodes and pixel nodes respectively. The variable $X_i$ can take a configuration $\mathbf{x}_i$ from its candidate space $\mathcal{U}_i$, while the variable $Y_i$ can take a value $y_i$ from label space $\mathcal{L}$. We define the pixel label space[1] $\mathcal{L} = \{0, \cdots m\}$, where $m$ is the number of triangle parts produced by the clique set $\mathcal{A}$ as defined in the last section. Given an image $\mathbf{I}$, the segmentation problem is formulated as an estimation of an optimal configuration $(\mathbf{X} = (\mathbf{x}_i)_{i\in\mathcal{V}_m}, \mathbf{Y} = (y_i)_{i\in\mathcal{V}_p})$ of all the nodes over model space $\mathcal{U} = \prod_{i\in\mathcal{V}_m} \mathcal{U}_i$ and labeling space $\mathcal{L}^k$.

$$\begin{aligned} (\mathbf{X},\mathbf{Y})^{\text{opt}} &= \arg \min_{\mathbf{X}\in\mathcal{U},\mathbf{Y}\in\mathcal{L}^k} E(\mathbf{X},\mathbf{Y},\mathbf{I}) \\ E(\mathbf{X},\mathbf{Y},\mathbf{I}) &= E^{(1)}(\mathbf{X},\mathbf{I}) + E^{(2)}(\mathbf{Y},\mathbf{I}) + E^{(3)}(\mathbf{X},\mathbf{Y}) \end{aligned} \tag{3}$$

where the MRF energy $E(\mathbf{X},\mathbf{Y},\mathbf{I})$ contains model-based energy $E^{(1)}$, pixel-based energy $E^{(2)}$ and interaction-based energy $E^{(3)}$.

### 3.2 Model-Based Energy

This energy is composed by data-related term and prior term [14]. The data term encodes the image likelihood given a model configuration. The prior term encodes spatial constraints of a model configuration with respect to the shape prior manifold. We formulate the model-based energy $E^{(1)}(\mathbf{X},\mathbf{I})$ as follows.

$$E^{(1)}(\mathbf{X},\mathbf{I}) = \lambda_1 \cdot \sum_{a\in\mathcal{A}} \Phi^{(1)}(\mathbf{x}_a) + \lambda_2 \cdot \sum_{c\in\mathcal{C}} \Psi^{(1)}(\mathbf{x}_c) \tag{4}$$

where $\lambda_1, \lambda_2$ are the weights of data term and prior term respectively.

The data term captures the homogeneity properties of the object region inside the model. Based on the model triangulation, the regional term can be factorized into higher order potentials $\Phi^{(1)}$ on data triplet set $\mathcal{A}$ introduced in section 2.

$$\Phi^{(1)}(\mathbf{x}_a) = \iint_{\mathbf{x}_a} \mathbf{L} ds, \;\; a \in \mathcal{A} \tag{5}$$

[1] When a pixel takes a non-zero value, it is labeled as a triangle part of the object, otherwise it is labeled as background.

where data potential $\varPhi^{(1)}$ encodes the integral of image likelihood function $\mathbf{L}$ over the triangle area $\mathbf{x}_a$. We denote likelihood[2] $\mathbf{L}_i = \log \frac{p_{\text{bck}}(\mathbf{I}_i)}{p_{\text{obj}}(\mathbf{I}_i)}$.

The prior term is formulated with the prior probability $p(\mathbf{X})$ defined in Eq.(2). It is factorized into potentials $\varPsi^{(1)}$ defined on prior clique set $\mathcal{C}$.

$$\varPsi^{(1)}(\mathbf{x}_c) = -\log p_c(\mathbf{x}_c),\ c \in \mathcal{C} \tag{6}$$

where the distribution probabilities $p_c$ are learned from training. We generalize the model-based interactions $\mathcal{C}_m = \{\mathcal{A}, \mathcal{C}\}$, where both subsets are triplet cliques.

### 3.3 Pixel-Based Energy

The energy $E^{(2)}$ also consists of a data term and a prior term as in [2].

$$E^{(2)}(\mathbf{Y}, \mathbf{I}) = \lambda_3 \cdot \sum_{i \in \mathcal{V}_p} \varPhi^{(2)}(y_i) + \lambda_4 \cdot \sum_{(i,j) \in \mathcal{C}_p} \varPsi^{(2)}(y_i, y_j) \tag{7}$$

where $\lambda_3, \lambda_4$ are the weights. The data term encodes the image likelihood over the pixel assignments. The unary potential $\varPhi^{(2)}$ encodes the individual penalties for labeling pixel $i$ as object or background.

$$\varPhi^{(2)}(y_i) = \begin{cases} -\log p_{\text{bck}}(\mathbf{I}_i) & \text{if } y_i = 0 \\ -\log p_{\text{obj}}(\mathbf{I}_i) & \text{otherwise} \end{cases} \tag{8}$$

where label $y_i = 0$ assigns the pixel $i$ as background, otherwise non-zero value assigns the pixel $i$ as object. We denote $p_{\text{bck}}$ and $p_{\text{obj}}$ in footnote 2. The prior term penalizes the inconsistency of the pixel labels within a neighborhood system (*e.g.* 8-connected) which is defined by pairwise clique set $\mathcal{C}_p$.

$$\varPsi^{(2)}(y_i, y_j) = \begin{cases} 0 & \text{if } y_i = y_j \\ W & \text{otherwise} \end{cases} \tag{9}$$

where pairwise potential $\varPsi^{(2)}$ constraints the neighboring pixel $i$ and pixel $j$ to have the same label, and $W$ is a penalizing parameter.

### 3.4 Interaction-Based Energy

The interaction energy is the key of producing consistency between model space and labeling space. This consistency is held when given a shape configuration $\mathbf{X}$, pixels inside/close to the shape boundary should be more likely labeled as object. Due to its dependency on the global shape, it is difficult to be encoded in the framework where both model solution and pixel labeling are sought for at the same time. Using the model triangulation, the model-pixel interaction can be factorized into constraints between pixel and model parts (triplets).

$$E^{(3)}(\mathbf{X}, \mathbf{Y}) = \sum_{(i,a) \in \mathcal{C}_{int}} \varPhi^{(3)}(y_i, \mathbf{x}_a),\quad \varPhi^{(3)}(y_i, \mathbf{x}_a) = -\log p(y_i | \mathbf{x}_a) \tag{10}$$

[2] $p_{\text{obj}}, p_{\text{bck}}$ are the appearance distribution models of object and background obtained from a training set.

The potential $\Phi^{(3)}$ is a third-order term, encoding the dependency of a pixel label and a regional triplet. The interaction clique set $\mathcal{C}_{int} = \{(i,a)|i \in \mathcal{V}_p, a \in \mathcal{A}\}$ connects every pixel with every regional triangle. The distribution $p(y_i|\mathbf{x}_a)$ of the pixel label conditioned on a triplet, uses a softmax function.

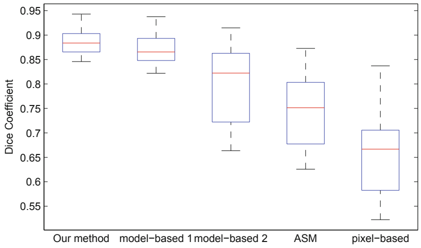
$$\begin{cases} P(y_i = l(a)|\mathbf{x}_a) = \left[1 + \exp(dist(i, \mathbf{x}_a))\right]^{-1} \\ P(y_i \neq l(a)|\mathbf{x}_a) = \left[1 - p(y_i = l(a)|\mathbf{x}_a)\right]/m \end{cases} \tag{11}$$

where label value $l(a) \in \{1, \cdots, m\}$ equals to the index of triplet $a$; $m$ is the number of the regional triplets in clique $\mathcal{A}$. We denote $dist(i, \mathbf{x}_a)$ a signed Chamfer distance of the pixel $i$ to the triangle $\mathbf{x}_a$ boundary (*i.e.* negative/positive when pixel is inside/outside the triangle). Hence, all energy terms are defined in MRF formulation and we can use TRW-S algorithms [8] for MAP-MRF inference. To search for an optimal model configuration, we use the same strategy as in [1].

## 4 Experimental Validation

A dataset of 40 2D tagged cardiac MR images (100*100 pixels) is used to validate our method. The ground truth provided by experts is used for both training and validation. Gabor features are used as the image representation to deal with tagged MR images. We performed a leave-one-out cross validation on the whole dataset. The experiments were run on a 2.8GHz, 12GB Ram computer and our segmentation took a couple of seconds per image.

Some visual results of two test images are presented in Fig. 4. The first column is our results in both model space $\mathbf{X}$ (yellow contours) and label space $\mathbf{Y}$. The second column shows the results from model-based module (using only energy $E^{(1)}$) [14] (blue contours) and pixel-based module (using only energy $E^{(2)}$) respectively. The third column provides the ground truth (green contours) and the pixel-wise difference image between our labeling result and the ground truth, where the gray pixels are correct labeled, the white/black pixels are wrongly labeled as object/background. The fourth column zooms in the area inside the red box (shown in the third column) with our model results (yellow contours) and model results by [14] (blue contours). We can see that only pixel-based method is sensitive to the complicated background and noise. The only model-based results are globally correct, but do not give accurate segmentation around the boundary locally (*e.g.* see the zoom in effects in the fourth column). The results from our



| Method | mean | std |
|---|---|---|
| our method | 0.8882 | 0.0263 |
| model-based 1 | 0.8709 | 0.0298 |
| model-based 2 | 0.7966 | 0.0761 |
| ASM | 0.7439 | 0.0710 |
| pixel-based | 0.6560 | 0.0781 |

**Fig. 3.** Comparisons on dice coefficients. Left: box plot. Right: statistic figures.

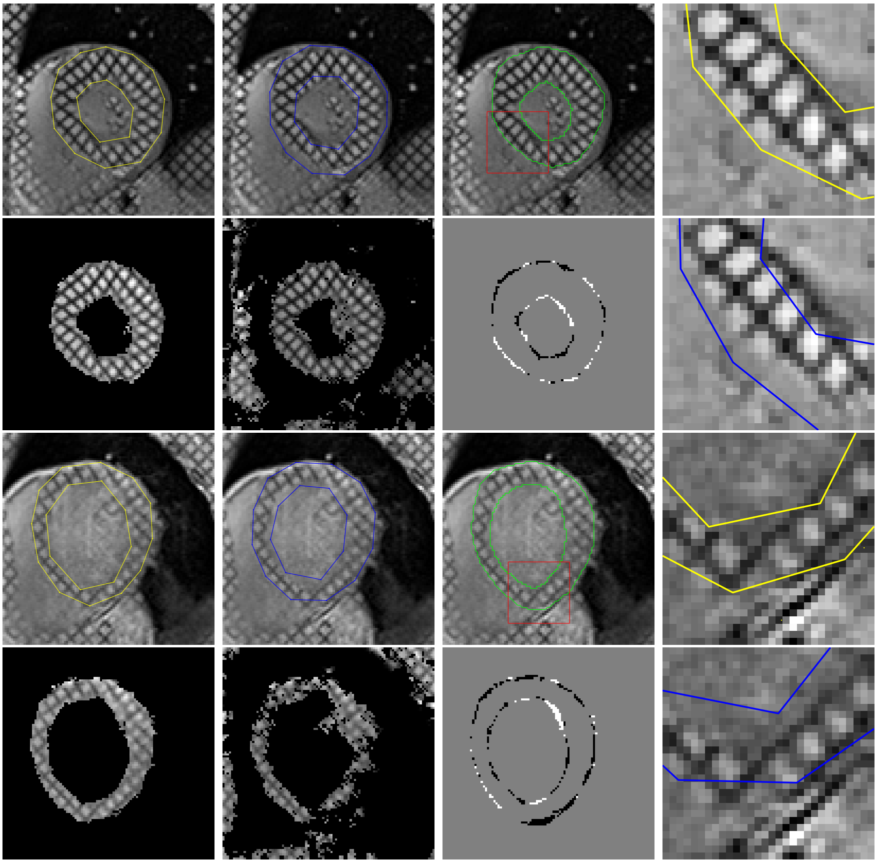


**Fig. 4.** Segmentation results of 2 test images. The columns from left to right are our results, only model/pixel based results, ground truth and comparison, zoom effects.

integrated framework can overcome this defect, showing flexibility to get local deviations as well as producing pixel-wise labeling result at the same time. Our method also deals well with the varying scales of the object (*e.g.* the scale of the inner contour shows large variability) thanks to the pose-invariant shape prior.

For both quantitative evaluation and comparison purposes, we present in Fig.3 the Dice coefficient distributions obtained respectively by (1) our hybrid method, (2) model-based method 1 [14], (3) model-based method 2 [1], (4) standard ASM method and (5) pixel-base method. Noted that a higher Dice coefficient implies a better segmentation result, Fig. 3 highlights the better performance of our method compared with the previous methods.

## 5 Conclusions

In this paper we propose a novel approach to address jointly model/image-based segmentation using a higher order graphical model. The proposed formulation can easily encode regional support, meanwhile being able to account for shape variability unseen during training. Furthermore, it produces states of the art results in particular when exact boundary delineation is of interest through the combined model-pixel graph. To the best of our knowledge, this is the first method that recovers a consistent solution between the model and the image space in a single shot optimization framework, while being pose-invariant.

The formulation involves more weight coefficients of different energy terms that need to be tuned to optimize the result. This defect can be addressed by estimating the parameters through a MRF training algorithm which can deal with a larger number of parameters. It is also necessary to extend the method in 3D cases where numerous challenging segmentation problems do exist in particular in medical image analysis. Compact modeling of shape variability is critical in terms of complexity (number of higher order cliques). Last but not least, understanding the varying importance of these cliques with respect to modeling of shape variations is also important. Recent progress on MRF learning could be a natural path towards adjusting the local contributions of the model.

## References

1. Besbes, A., Komodakis, N., Langs, G., Paragios, N.: Shape priors and discrete mrfs for knowledge-based segmentation. In: CVPR, pp. 1295–1302 (2009)
2. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient n-d image segmentation. IJCV 70, 109–131 (2006)
3. Bray, M., Kohli, P., Torr, P.: PoseCut: Simultaneous segmentation and 3D pose estimation of humans using dynamic graph-cuts. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 642–655. Springer, Heidelberg (2006)
4. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. PAMI 24, 603–619 (2002)
5. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. PAMI 23(6), 681–685 (2001)
6. Cootes, T., Taylor, C., Cooper, D., Graham, J., et al.: Active shape models-their training and application. CVIU 61, 38–59 (1995)
7. Glocker, B., Komodakis, N., Paragios, N., Glaser, C., Tziritas, G., Navab, N.: Primal/dual linear programming and statistical atlases for cartilage segmentation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 536–543. Springer, Heidelberg (2007)
8. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. PAMI 28(10), 1568–1583 (2006)
9. Shi, J., Malik, J.: Normalized cuts and image segmentation. PAMI 22 (2000)
10. Staib, L., Duncan, J.: Boundary finding with parametrically deformable models. PAMI 14(11), 1061–1075 (1992)
11. Taron, M., Paragios, N., Jolly, M.: Registration with uncertainties and statistical modeling of shapes with variable metric kernels. PAMI 31(1), 99–113 (2009)
12. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the mumford and shah model. IJCV 50, 271–293 (2002)
13. Wang, C., de La Gorce, M., Paragios, N.: Segmentation, ordering and multi-object tracking using graphical models. In: ICCV, pp. 747–754 (2009)
14. Xiang, B., Wang, C., Deux, J., Rahmouni, A., Paragios, N.: 3d cardiac segmentation with pose-invariant higher-order mrfs. In: ISBI, pp. 1425–1428 (2012)