

Model-Based Segmentation Using Graph Representations

D. Seghers, J. Hermans, D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens

Katholieke Universiteit Leuven, Faculties of Medicine and Engineering, Medical Image Computing (Radiology - ESAT/PSI), University Hospital Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium
`dieter.seghers@uz.kuleuven.ac.be`

Abstract. A generic supervised segmentation approach is presented. The object is described as a graph where the vertices correspond to landmark points and the edges define the landmark relations. Instead of building one single global shape model, a priori shape information is represented as a concatenation of local shape models that consider only local dependencies between connected landmarks. The objective function is obtained from a maximum a posteriori criterion and is build up of localized energies of both shape and landmark intensity information. The optimization problem is discretized by searching candidates for each landmark using individual landmark intensity descriptors. The discrete optimization problem is then solved using mean field annealing or dynamic programming techniques. The algorithm is validated for hand bone segmentation from RX datasets and for 3D liver segmentation from contrast enhanced CT images.

1 Introduction

Supervised segmentation has become very popular in the medical image processing domain due to its generic nature, robustness and high accuracy. In general these methods extract shape and gray-value appearance information from a set of training images and corresponding segmentations. A very popular approach is to build a global model of the shape as for example proposed by [1]. Unfortunately, these global models do not fit accurately to unseen images, especially in the case of a low number of training shapes. A way to cope with this is the use of more flexible models, either by superimposing local degrees of freedom onto a global shape description as for example the work of Heimann *et al.* [2], or by limiting the shape model to local descriptors as for example [3, 4]. In a previous contribution [5] the segmentation problem was formulated as a minimal cost path problem (MCP) where intensity costs are assigned to the nodes (landmarks) and shape costs are assigned to node-to-node transitions. A drawback of this method is that the shapes are limited to contours. The objective of this work is to generalize the MCP method to a more generic graph-based object representation.

2 Method

2.1 Training Set

The training set consists of a set of images and a set of corresponding segmentations. The segmented object is mathematically represented as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: the vertices $\mathcal{V} = \{1, \dots, n\}$ correspond to the landmark points $\mathbf{l}_1, \dots, \mathbf{l}_n$ and the edges \mathcal{E} contain unordered pairs of distinct vertices $\{i, j\}$. The number of edges will be denoted with t . An alternative way to formulate the interrelationship between the landmarks is by defining the neighborhood system $\mathcal{N} = \{\mathcal{N}_i | \forall i \in \mathcal{V}\}$ where \mathcal{N}_i is the set of vertices neighboring i . The presented approach assumes that the images in the training set are aligned affinely, either by acquisition or by registration.

The behavior of the object will be described statistically by building a model of the gray-level appearance in the image around each landmark and by modeling the shape.

2.2 Intensity Model

The intensity model is constructed by estimating the probability distribution functions (PDF) of the intensity patterns \mathbf{f}_i in the image around each landmark i individually. First, an intensity descriptor $\mathbf{f}_i = F(I, \mathbf{l}_i)$ is defined that extracts the pattern \mathbf{f}_i around the location \mathbf{l}_i in the image I .

In the learning phase, the pattern \mathbf{f}_i is extracted from each training image and at each landmark \mathbf{l}_i . From these samples, the PDF $p(\mathbf{f}_i)$ is estimated for each landmark individually. Writing the PDF as a Gibbs function

$$p(\mathbf{f}_i) = \frac{1}{z_i} \exp\left(-\frac{1}{2}d_i(\mathbf{f}_i)\right) \quad (1)$$

allows to work with energies $d_i(\mathbf{f}_i)$ instead of probabilities $p(\mathbf{f}_i)$. A location in a test image with a low intensity energy has an intensity pattern which is similar to what is found in the training images and is likely to coincide with the true landmark location from an intensity point of view.

When the PDF is modeled as a Gaussian distribution, the energy $d_i(\mathbf{f}_i)$ becomes the mahalanobis distance between \mathbf{f}_i and its mean:

$$d_i(\mathbf{f}_i) = (\mathbf{f}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{f}_i - \boldsymbol{\mu}_i) \quad (2)$$

with $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$, respectively the mean and covariance of \mathbf{f}_i .

The intensity features of different landmarks are modeled independently, hence the joint PDF of $\mathbf{f} = (\mathbf{f}_1, \dots, \mathbf{f}_n)$ becomes

$$p(\mathbf{f}) = \frac{1}{Z_f} \exp\left(-\frac{1}{2}E_f(\mathbf{f})\right) \quad (3)$$

with $E_f(\mathbf{f}) = \sum_{i=1}^n d_i(\mathbf{f}_i)$ the global intensity energy and Z_f a normalizing constant.

The same approach as in [5] is adopted to define the intensity descriptor $F(I, \mathbf{l}_i)$. First, a set of N feature images is computed from the original image I . Secondly, intensity profiles $\mathbf{f}_i^{(k)}$ are extracted from the feature images $I^{(k)}$ by taking samples on a sphere centered at \mathbf{l}_i . Alternatively, linear instead of spherical profiles can be used by taking samples along the image gradient in \mathbf{l}_i . The intensity pattern $\mathbf{f}_i = F(I, \mathbf{l}_i)$ is then constructed by concatenating the N feature profiles $\mathbf{f}_i^{(k)}$ resulting in one large intensity pattern \mathbf{f}_i . The PDF $p(\mathbf{f}_i)$ (Eq. 1) is estimated by modeling each feature vector $\mathbf{f}_i^{(k)}$ individually. In the case of a Gaussian, the covariance Σ_i in the energy (Eq. 2) becomes a block diagonal matrix as the dependencies between different features are ignored. Consequently, the computed energy is an overestimation of the true value.

2.3 Shape Model

The shape is described by the positions of the landmarks $\mathbf{l} = \mathbf{l}_V = (\mathbf{l}_1, \dots, \mathbf{l}_n)$. Similar as for the intensity patterns, a model of the shape is built by estimating the joint shape PDF $p(\mathbf{l})$ from a set of training shapes. The model relies on two assumptions. First, the probability function of an edge vector does not depend on its location in space: $p(\mathbf{l}_j - \mathbf{l}_i | \mathbf{l}_i) = p(\mathbf{l}_j - \mathbf{l}_i)$ with $j \in \mathcal{N}_i$, implying $p(\mathbf{l}_j | \mathbf{l}_i) = p(\mathbf{l}_j - \mathbf{l}_i)$. This assumption also implies that the shape PDF is invariant for translations: $p(\mathbf{l}_i) = c^{te}$. A second model supposition (*markovianity*) is that a landmark i only interacts with its neighbors: $p(\mathbf{l}_i | \mathbf{l}_{V - \{i\}}) = p(\mathbf{l}_i | \mathcal{N}_i)$. Hence, the shape obeys the definition of a markov random field (MRF). According to the Hammersley-Clifford theorem [6] the joint PDF of an MRF can be written as

$$p(\mathbf{l}) = \frac{1}{Z_1} \exp\left(-\frac{1}{2} E_1(\mathbf{l})\right) \tag{4}$$

The energy $E_1(\mathbf{l})$ is a sum of local energy functions over all possible cliques¹ of the graph and Z_1 is a normalizing constant. First, the shape energy is computed for a trivial graph consisting of only one triangle: $\mathcal{V} = \{1, 2, 3\}$ and $\mathcal{E} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$. The joint PDF written in the energy space becomes

$$\log p(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) = \log p(\mathbf{l}_1) + \log p(\mathbf{l}_2 | \mathbf{l}_1) + \log p(\mathbf{l}_3 | \mathbf{l}_1, \mathbf{l}_2) \tag{5}$$

If only the influence of the edges $\{1, 2\}$ and $\{1, 3\}$ is considered, the log equation (Eq. 5) can be approximated as

$$\log p(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) \approx \log p(\mathbf{l}_1) + \log p(\mathbf{l}_2 | \mathbf{l}_1) + \log p(\mathbf{l}_3 | \mathbf{l}_1) \tag{6}$$

In order to obtain an approximation that takes all edges into account, Eq. 6 is considered for all three possible combinations. Averaging them leads to

$$\log p(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) \approx c^{te} + \frac{2}{3} (\log p(\mathbf{l}_2 - \mathbf{l}_1) + \log p(\mathbf{l}_3 - \mathbf{l}_1) + \log p(\mathbf{l}_3 - \mathbf{l}_2)) \tag{7}$$

¹ A clique is a subset of a graph such that every two vertices in the clique are neighbors.

Similar as in Eq. 1, the PDF of an edge $\mathbf{l}_j - \mathbf{l}_i$ can be written using the *edge energy* $d_{ij}(\mathbf{l}_i, \mathbf{l}_j)$ expressing the likelihood of the edge $\mathbf{l}_j - \mathbf{l}_i$. When the edge distribution is modeled as a Gaussian, the energy $d_{ij}(\mathbf{l}_i, \mathbf{l}_j)$ is the mahalanobis distance between $\mathbf{l}_j - \mathbf{l}_i$ and its expected value. Remark that the edge energy is symmetrical, i.e. $d_{ij}(\mathbf{l}_i, \mathbf{l}_j) = d_{ji}(\mathbf{l}_j, \mathbf{l}_i)$. From Eq. 4 and Eq. 7, the triangle shape energy becomes

$$E_1(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) = \frac{2}{3} (d_{12}(\mathbf{l}_1, \mathbf{l}_2) + d_{13}(\mathbf{l}_1, \mathbf{l}_3) + d_{23}(\mathbf{l}_2, \mathbf{l}_3)) \quad (8)$$

For a general graph with n vertices and t edges, a similar expression as Eq. 8 can be derived

$$E_1(\mathbf{l}) = \frac{n-1}{t} \sum_{\{i,j\} \in \mathcal{E}} d_{ij}(\mathbf{l}_i, \mathbf{l}_j) \quad (9)$$

2.4 Objective Function

Criterion. The segmentation problem is formulated using a maximum a posteriori (MAP) probability criterion:

$$\mathbf{I}^* = \arg \max_{\mathbf{I}} p(\mathbf{I}|I) = \arg \max_{\mathbf{I}} \frac{p(I|\mathbf{l})p(\mathbf{l})}{p(I)} \quad (10)$$

The shape prior $p(\mathbf{l})$ is given by Eq. 4 and Eq. 9. The probability $p(I|\mathbf{l})$ is proportional to $p(\mathbf{f})$ as given by Eq. 3 with $\mathbf{f} = F(I, \mathbf{l})$. Hence, the criterion (Eq. 10) becomes

$$\mathbf{I}^* = \arg \min_{\mathbf{I}} (E_f(F(I, \mathbf{l})) + E_1(\mathbf{l})) \quad (11)$$

$$= \arg \min_{\mathbf{I}} \left(\sum_{i=1}^n d_i(F(I, \mathbf{l}_i)) + \gamma \sum_{\{i,j\} \in \mathcal{E}} d_{ij}(\mathbf{l}_i, \mathbf{l}_j) \right) \quad (12)$$

with $\gamma = (n-1)/t$.

Discretization. As the objective function (Eq. 12) has many local optima it is difficult to optimize in the continuous domain. Therefore, the segmentation task is converted to a discrete labeling problem. For each landmark i , a set of candidate locations is obtained by evaluating the intensity energy $d_i(F(I, \mathbf{l}))$ at each point \mathbf{l} of a search grid and selecting the m lowest-energy locations. The result is a set of candidates $\mathcal{C} = \{\{\mathbf{l}_{ik}, \mathbf{f}_{ik}\}_{k=1}^m\}_{i=1}^n$. At this point, the segmentation problem becomes a labeling problem $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ where one candidate per landmark needs to be selected. \mathbf{x}_i represents the choice for landmark i : $\sum_{k=1}^m x_{ik} = 1$ and $x_{ik} = 1$ if candidate k is selected. The PDF $p(\mathbf{l}|I)$ is now converted into a discrete probability distribution $P(\mathbf{x}|\mathcal{C})$ with energy

$$E(\mathbf{x}) = \sum_{i=1}^n \sum_{a=1}^m (x_{ia} d_i(\mathbf{f}_{ia})) + \frac{\gamma}{2} \sum_{i=1}^n \sum_{a=1}^m \left(x_{ia} \sum_{j \in \mathcal{N}_i} \sum_{b=1}^m x_{jb} d_{ij}(\mathbf{l}_{ia}, \mathbf{l}_{jb}) \right) \quad (13)$$

Hence, an intensity energy is assigned to each selected candidate landmark (first term in Eq 13), and a local shape energy is assigned to each selected edge. The theoretical value for γ is $(n - 1)/t$. In the implementation of this method, the value for γ is also used to compensate for the overestimated intensity energy (Eq. 2). The optimal γ -value is obtained from a leave-one-out strategy during the training phase yielding a value always larger than $(n - 1)/t$.

2.5 Optimization

Dynamic Programming (DP). In the case when the graph is an open contour with n landmarks and $n - 1$ edges, the problem is reduced to a minimal cost path problem which can be solved using dynamic programming [6, 5] with computational complexity $O(nm^2)$. Also more complex graphs can be handled using DP, but with increasing complexity. The computational complexity can be found out by eliminating the graph landmark by landmark. The computational complexity is $O(nm^{a+1})$ with a the number of edges that disappear each time a landmark is eliminated.

Iterative DP. From a computational point of view, DP is not acceptable for complex graphs. An approximating iterative procedure is proposed instead. Prior to optimization, the graph is divided into a number of (overlapping) subgraphs (paths). The paths are generated randomly, starting from a specific edge and iteratively adding concatenating edges as long as the path is not self-intersecting. From every edge in \mathcal{E} a random path is started resulting into a total of t subgraphs $\mathcal{V}^{(i)}$. Once all the paths are generated, an iterative optimization procedure is started as follows. Each subgraph $\mathcal{V}^{(i)}$ is optimized globally using DP and the result yields votes for the selected landmark candidates. After optimizing every random path, the least voted candidates are removed. This procedure is repeated till only one candidate is left for every landmark.

Mean field annealing. The discrete labeling \mathbf{x} satisfies the condition of an MRF and listens to the Gibbs distribution

$$P(\mathbf{x}|\mathcal{C}) = \frac{1}{Z_{\mathbf{x}}} \exp\left(-\frac{1}{T}E(\mathbf{x})\right) \tag{14}$$

with $T = 2$ and with $E(\mathbf{x})$ computed as in Eq. 13. According to Eq. 14, the temperature T can be altered without influencing the optimum \mathbf{x}^* . At the other hand, the mean $\bar{\mathbf{x}}$ changes with T as follows:

$$\lim_{T \rightarrow 0+} \bar{\mathbf{x}}_T = \lim_{T \rightarrow 0+} \sum_{\mathbf{x}} \mathbf{x}P(\mathbf{x}) = \mathbf{x}^* \tag{15}$$

The rationale behind mean field annealing [6] is that instead of minimizing the energy $E(\mathbf{x})$ directly, one can try to estimate the mean field $\bar{\mathbf{x}}_T$ at a sufficiently high temperature and then track it down as the temperature is lowered towards

zero. The estimated mean field β is computed from the following set of nonlinear equations:

$$\alpha_{ik} = \exp \left(-\frac{1}{T} \left(d_i(\mathbf{f}_{ik}) + \gamma \sum_{j \in \mathcal{N}_i} \sum_{b=1}^m \beta_{jb} d_{ij}(\mathbf{l}_{ik}, \mathbf{l}_{jb}) \right) \right) \quad (16)$$

$$\beta_{ik} = \frac{\alpha_{ik}}{\sum_{a=1}^m \alpha_{ia}} \quad (17)$$

$\forall i = 1, \dots, n$ and $\forall k = 1, \dots, m$. The set is solved using iterative resubstitution: The values for $\beta_{ik}^{(i)}$ are computed from $\alpha_{ik}^{(i)}$ using Eq. 17. The updated values $\alpha_{ik}^{(i+1)}$ are computed with Eq. 16. This procedure is started from an initial point $\alpha_{ik}^{(0)}$ and iterated till the values for α_{ik} and β_{ik} are sufficiently stable. The mean field equations are solved several times, starting at a sufficiently high T , and gradually decreasing the temperature towards zero resulting in a mean β that converges to a binary assignment.

3 Experiments and Results

The algorithm is validated for two applications: segmentation of hand bones from RX and liver segmentation from contrast enhanced CT scans.

3.1 Hand Bones

The segmentation task consists of the delineation of 11 bones in the hand (bones of the thumb, the small and middle finger) from RX datasets. A hierarchical segmentation strategy is illustrated in Fig. 1. In a first step, the algorithm of Sect. 2 is applied with a model trained from graphs consisting of 11 vertices positioned at the centers of the hand bones as shown in Fig. 1(a). Applying a segmentation with this model enables the localization of each bone individually. Fig. 1(b) shows an unseen hand radiograph to be segmented. The intensity PDF's (Eq. 1) of landmarks 1, 4 and 11 are superimposed on the image in the corresponding search regions. The search region of \mathbf{l}_i is chosen such that each \mathbf{l}_i of the training shapes is covered. In a second step, eleven individual contour models consisting of 20 to 40 landmarks (Fig. 1(c)) are used to delineate each bone individually. The segmentation is performed on sub-images centered around the detected centers from the first step. Fig. 1(d) shows such an extracted image together with the intensity PDF's of landmarks 8 and 18. The delineation of bone 11 is shown in Fig. 1(e).

This hierarchical segmentation setup is validated on a set of 50 training images (width of 256 pixels) of both adults and children using a leave-one-out approach. The intensity model was built using spherical profiles with a radius of 8 pixels for the first model and profiles with a radius of 1 pixel to delineate the individual bones. Optimizing with iterative DP resulted in a mean overlap and mean distance error of 87.5% and 1.36 pixels respectively. Using mean field annealing scored slightly worse with an overlap of 86.2% and 1.46 pixels distance error.

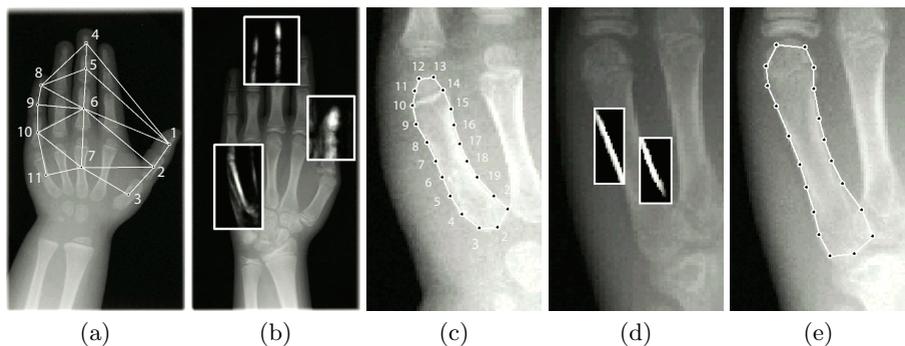


Fig. 1. Hand segmentation strategy: (a) model for bone localization; (b) intensity PDF of landmark 1, 4 and 11; (c) model for bone delineation; (d) intensity PDF of landmark 8 and 18; (e) automated bone delineation

3.2 Liver

A set of 20 contrast enhanced CT datasets and corresponding segmentations of the liver were obtained from [7]. Prior to segmentation, an affine registration step was applied to align the images. The liver was described as a surface mesh consisting of $n = 3277$ landmarks and $t = 9825$ edges. Segmentation is carried out with a multi-resolution-approach starting from large profiles (large radius) and large search regions and ending with smaller and more accurate profiles in smaller search regions. Three resolution stages were carried out with linear profiles sampled along the image gradient with sizes ranging between 30 and 6 millimeters. In between consecutive resolution stages, a postprocessing step was needed to counter the following problem. Prior to optimization, candidate locations are searched for every landmark. If no proper locations were generated for a particular landmark due to a failing intensity descriptor, the segmented object contained outliers. To overcome this, the outlier locations with excessive shape energies were relaxed to lower shape energy locations.

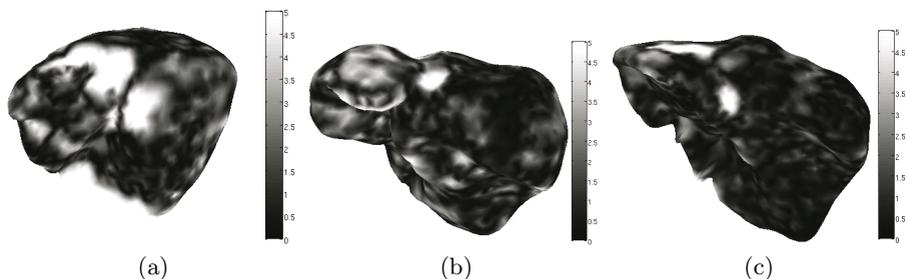


Fig. 2. Surface distance errors for (a) the worst (3.5mm), (b) an average (1.4mm) and (c) the best (0.9mm) segmentation. Black and white correspond to 0 and 5mm respectively.

The segmentation scheme was validated using leave-one-out on the training set and applying the scoring system as in [7] which is based on five metrics (volumetric overlap, volume difference, surface distance, root mean square surface distance and maximum surface distance). For comparison, a score of 75% implies a performance as good as human. The best results were obtained with iterative DP with an mean score of 67% compared to 57% for mean field annealing. The mean overlap and closest distance error was 91.7% and 1.47mm respectively. Fig. 2 displays the distance errors of the worst, middle and best case with scores of 33%, 72% and 83% respectively.

4 Discussion

A supervised image segmentation method based on graph representations is proposed. The model is built from local statistical models of both shape and gray-level appearance resulting in an approach that accurately fits to unseen data. The discretization of the objective function obtained from a MAP criterion, allows to formulate the segmentation task as a labeling problem and thus enables robust optimization techniques as dynamic programming and mean field annealing. The generic nature of the method is illustrated for hand bone delineation from RX datasets and for liver segmentation from contrast enhanced CT scans. Despite the heuristic nature of the iterative DP method, better results were obtained than optimizing with mean field annealing.

References

- [1] Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models - their training and applications. *Computer Vision and Image Understanding* 61(1), 38–59 (1995)
- [2] Heimann, T., Wolf, I., Meinzer, H.: Active shape models for a fully automated 3D segmentation of the liver – an evaluation on clinical data. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) *MICCAI 2006*. LNCS, vol. 4191, pp. 41–48. Springer, Heidelberg (2006)
- [3] Amit, Y., Kong, A.: Graphical templates for model registration. *IEEE Trans. Pattern Anal. Machine Intell.* 18(3), 225–236 (1996)
- [4] Felzenszwalb, P., Huttonlocker, D.: Pictorial structures for object recognition. *Int. J. Comput. Vis.* 61(1), 55–79 (2005)
- [5] Seghers, D., Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P.: Minimal intensity and shape cost path segmentation. *IEEE Trans. Med. Imag.* 26(8), 1115–1129 (2007)
- [6] Li, S.: *Markov Random Field Modeling in Computer Vision*. Springer, Heidelberg (1995)
- [7] van Ginneken, B., Heimann, T., Styner, M.: 3D segmentation in the clinic: A grand challenge. In: Heimann, T., Styner, M., van Ginneken, B. (eds.) *3D Segmentation in the Clinic: A Grand Challenge*, pp. 7–15 (2007)