Image Segmentation With a Shape Prior Based on Simplified Skeleton

Boris Yangel, Dmitry Vetrov

Lomonosov Moscow State University

Abstract. In the paper we propose a new deformable shape model that is based on simplified skeleton graph. Such shape model allows to account for different shape variations and to introduce global constraints like known orientation or scale of the object. We combine the model with low-level image segmentation techniques based on Markov random fields and derive an approximate algorithm for the minimization of the energy function by performing stochastic coordinate descent. Experiments on two different sets of images confirm that usage of proposed shape model as a prior leads to improved segmentation quality.

Key words: image segmentation, shape prior, MRF, skeleton

1 Introduction

Image segmentation is an important, but inherently ambiguous problem. Practical segmentation systems rely on some user input and provide a way to combine that input with the low-level cues, such as color distributions and contrast edges observed in the image. Sometimes high-level information about image like shape of segmented object is available. Intuitively it seems that the more prior information about segmented object is involved in the model, the better its final segmentation is. Bayesian framework provides an efficient way of combining both low- and high-level information within a unified framework. Low-level cues are usually taken into account by using well-examined Markov random fields (MRF) theory [1]. However the straightforward addition of high-level information to MRF framework makes it intractable hence some extensions of MRFs are needed.

Related work

Prior work on image segmentation with shape models can be divided into several classes. Some methods use shape prior in a form of hard object mask, which is represented either via level-sets [2] or via distance function [3]. To cope with missing information about object location in image, these methods use some iterative location re-estimation scheme combined with repeated segmentation of the image. While approaches based on a hard object mask can be quite robust when object shape is similar to mask, they are not applicable for classes of objects with high shape variability.

Similar approaches to shape modeling are used in [4] and [5]. In [4] shape prior is represented by a set of hard object masks. Branch-and-bound is used to choose right prior from that set. Paper [5] proposes usage of probability mask as a shape prior; it combines variational segmentation with a probability mask prior with a continuous analogue of branch-and-bound technique for object location estimation.

Another class includes approaches that limit the set of possible shapes of the object. Some examples include star-shape prior [6] and tightness prior [7]. These broad restrictions can be of great utility in particular situations, while completely useless in others. For example, mislabeled pixels can sometimes make object even more tight or star-shaped.

One more class of techniques that are close to our work in spirit includes approaches that represent shape as a set of rigid parts that may have various positions w.r.t. to one another. In [8] shape model is represented by a layered pictorial structure. Monte Carlo EM algorithm is used to obtain image labeling. Sampling from the posterior distribution on the parameters of pictorial structure is performed via loopy belief propagation. Similar approach is presented in [9], where stick figure of a human is used as a prior. In [10] shape model based on skeleton is used for part-based object detection.

Contribution

In this paper we propose new shape prior that represents object shape via simplified skeleton graph. Edges of the graph correspond to meaningful parts of an object. Radii assigned to vertices of the graph specify width of object parts. Not only can proposed shape model describe object shape variation and global shape constraints such as known orientation or scale, but it can also account for non-uniform scale of object parts via soft constraints on vertex radii.

We also propose a framework for combining MRF segmentation with a shape prior. Our framework leads to iterative segmentation process with two-step iterations. On the first step shape model is re-estimated via stochastic optimization. On the second step, MRF segmentation is combined with current shape model to produce new pixel labeling. We also show that proposed approach can be seen as a specific kind of EM algorithm.

2   An iterative approach to segmentation with a shape prior

In this section we propose an iterative approach to MRF segmentation with a shape prior based on a generative model. Then we discuss its relation to EM algorithm.

Our approach is based on a graphical model presented in figure 1. In this model random variable S represents shape model that produces pixel labels L_i (one per pixel). Labels L_i take values 0 (pixel belongs to background) and 1 (pixel belongs to object). Shape model does not produce pixel labels independently,
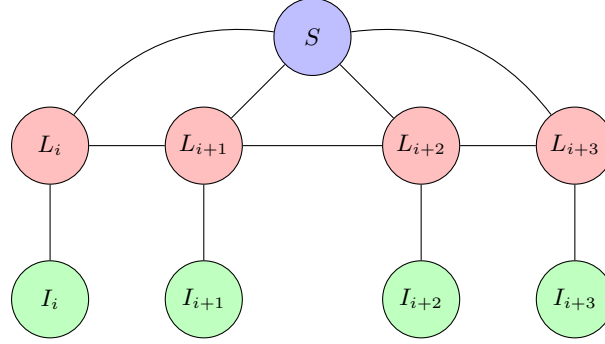
Fig. 1. Generative model for the segmentation with a shape prior

but in a way that neighboring pixels are likely to have the same label. This soft constraint is represented by edges between labels of the neighboring pixels. Finally, each image pixel I_i is independently generated by the label L_i using some class-specific color model.

2.1 Iterative segmentation as a coordinate descent

Joint probability in the discussed graphical model can be expressed as

$$P(S, L, I) = P(I \mid L)P(L \mid S)P(S) =$$
$$= \frac{1}{Z}f(S)\prod_i h_i(I_i, L_i)\prod_{(i,j)\in\mathcal{N}} \tilde{\phi}_{ij}(L_i, L_j, S), \tag{1}$$

where $\mathcal{N}$ is a neighborhood model and $\tilde{\phi}_{ij}$ are potentials corresponding to 3-cliques of the model graph. We further assume that each of the potentials $\tilde{\phi}_{ij}$ can be expressed as a product of the pairwise terms. Then we can finally rewrite joint probability as

$$P(S, L, I) = \frac{1}{Z}f(S)\prod_i h_i(I_i, L_i)\prod_{(i,j)\in\mathcal{N}} \phi_{ij}(L_i, L_j)\prod_i \phi_i(L_i, S). \tag{2}$$

Let us now state the problem of image segmentation as the problem of finding

$$\langle S^*, L^*\rangle = \arg\max_{S,L} P(S, L \mid I) = \arg\max_{S,L} P(S, L, I) =$$
$$= \arg\min_{S,L}\Big[-\log f(S) - \sum_i \big(\log h_i(I_i, L_i) + \log \phi_i(L_i, S)\big) - \tag{3}$$
$$- \sum_{(i,j)\in\mathcal{N}} \log \phi_{ij}(L_i, L_j)\Big].$$

Minimization of this expression can be performed by coordinate descent w.r.t. two groups of variables: L and S. In this case update expressions for each group

can be written as

$$S^{new} = \arg\min_{S} \Big[-\log f(S) - \sum_{i} \log \phi_i(L_i^{old}, S) \Big], \tag{4}$$

$$L^{new} = \arg\min_{L} \Big[-\sum_{i} \big(\log h_i(I_i, L_i) + \log \phi_i(L_i, S^{new}) \big) -$$
$$-\sum_{(i,j)\in\mathcal{N}} \log \phi_{ij}(L_i, L_j) \Big]. \tag{5}$$

Note that the update step for L is in fact a regular binary image segmentation problem with unary potentials modified by the shape prior. Thus, it can be efficiently solved using graph cuts for submodular pairwise terms ϕ_{ij}. On the other hand, update step for S is more challenging optimization problem. Optimization algorithm for it should be selected according to particular form of $f(S)$.

2.2   Relation to EM algorithm

Iterative segmentation approach discussed in section 2.1 can be seen as a particular form of expectation-maximization algorithm. If we consider S a group of latent variables and use EM approach to maximize posterior probability $P(L \mid I)$, M-step takes the form

$$L^{new} = \arg\max_{L} \Big[\mathbb{E}_{S \sim P^*(S)} \log P(I, S \mid L) + \log P(L) \Big] =$$
$$= \arg\max_{L} \Big[\mathbb{E}_{S \sim P^*(S)} \log P(I, L, S) \Big] =$$
$$= \arg\max_{L} \Big[\mathbb{E}_{S \sim P^*(S)} \sum_{i} \log \phi_i(L_i, S) + \tag{6}$$
$$+ \sum_{i} \log h_i(I_i, L_i) + \sum_{(i,j)\in\mathcal{N}} \log \phi_{ij}(L_i, L_j) \Big].$$

Distribution $P^*(S)$ comes from E-step and has form

$$P^*(S) = P(S \mid I, L^{old}) = \frac{P(S, I, L^{old})}{P(I, L^{old})} = \frac{1}{Z} \prod_{i} \phi_i(L_i^{old}, S) f(S). \tag{7}$$

It can be easily shown that the algorithm presented in section 2.1 is equivalent to approximating $P^*(S)$ with delta function centered at the distribution mode on the E-step. In this case E-step corresponds to an update of S, while M-step is equivalent to updating L by solving regular segmentation problem.

Interpretation of proposed segmentation algorithm as a particular case of EM can give rise to alternative ways of solving the problem. For example, instead of approximating $P^*(S)$ with a delta function, one could use Monte Carlo EM to approximate the expectation itself.

Fig. 2. Giraffe image with graph-based shape model.

3   Simplified figure skeleton as a shape prior

In this section we present a new shape prior that allows for controllable shape variation. We also discuss a way to build the prior into the segmentation approach presented in section 2.

3.1   Graph-based shape model

In order to handle significant shape variation, we propose a graph-based shape model where each edge encodes some meaningful part of the object. Radii assigned to each vertex of the graph allow us to encode variable width of each object part. This representation can be seen as a simplified version of object skeleton. One example of such a representation is shown in figure 2.

Soft constraints can be introduced into this shape model via MRF-like energy function

$$E(S) = -\log f(S) = \sum_i U_i(e_i) + \sum_{(i,j) \in \mathcal{N}_S} B_{ij}(e_i, e_j), \qquad (8)$$

where e_i denotes i-th edge of the shape graph and $\mathcal{N}_S$ is a set of all pairs of neighboring edge indices. Unary terms U_i in this case represent global constraints on the edge itself, like its scale or location. Binary terms B_{ij} can constrain
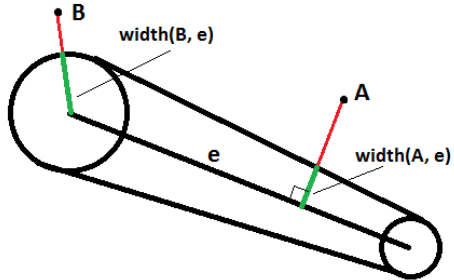
Fig. 3. Edge width and distance for points A and B.
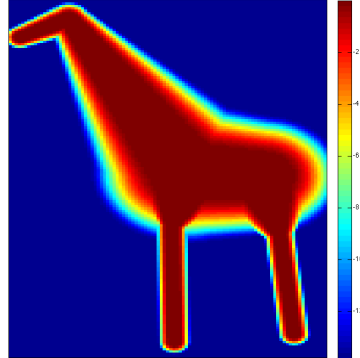
Fig. 4. $\log \phi_i(1, S)$ for the shape model of a giraffe.

relative sizes and angles between the connected object parts. This model can be easily made invariant to rigid transformations by removing all the global constraints and considering only relative ones. Parameters of the unary and binary terms can be learned from a set of labeled images using techniques like ML estimation. Particular forms of shape energy that we used in our experiments are covered in section 4.

3.2 Unary potentials

In order to complete description of the proposed shape prior, we should specify potential functions $\phi_i(L_i, S)$ from (3). It is natural to assume that pixels located near the edges of the shape graph will certainly belong to object, while pixels that are far from any edge will most likely belong to background. This observation yields the following expression for ϕ_i:

$$\phi_i(L_i, S) = L_i \max_j W(e_j, i) + (1 - L_i)(1 - \max_j W(e_j, i)). \qquad (9)$$

In this expression $W(e, i)$ denotes a function that decreases from 1 to 0 as the distance from the edge e to the pixel i increases. In our experiments we have used function W of the form

$$W(e, i) = \exp\left[-w \max\left\{0, \left(\frac{\mathrm{dist}(e, i) - \alpha \, \mathrm{width}(e, i)}{(1 - \alpha) \, \mathrm{width}(e, i)}\right)^p\right\}\right], \qquad (10)$$

where $\mathrm{dist}(e, i)$ is a distance from edge e to pixel i and $\mathrm{width}(e, i)$ is edge width for that pixel (see figure 3). This function holds 1 while distance from the edge goes from 0 to $\alpha \, \mathrm{width}(e, i)$, then it decreases in a way that for $\mathrm{dist}(e, i) = \mathrm{width}(e, i)$ it has value $\exp(-w)$. An example of unary potentials calculated by the proposed function can be seen in figure 4.

3.3 Shape fitting via simulated annealing

As it was said earlier, some segmentation methods use discrete [4] or continuous [5] versions of branch-and-bound method for shape fitting. We find out that local minima achieved with simulated annealing (SA) were good enough for the whole approach to work, so we decided to use it for S update step (4). On each SA iteration we slightly perturbed positions and radii of graph vertices to update current solution. Perturbation variance was proportional to temperature $T = \frac{1}{\log k}$ at iteration k. Optimization process usually converged in 2000-4000 iterations in our experiments. Annealing initialization is explained in section 4.2.

4 Experiments

Segmentation method presented in this paper was tested on two sets of images. One set was obtained by filtering giraffe photos used in [11], leaving only photos with giraffes in lateral view. Another set consisted of various images with capital "E" letter; it was built manually from various sources. Images from both sets lack reliable edge and color models for object and background, and therefore need some additional information like shape model to improve segmentation quality. Shape models for both giraffes and letters were set manually. Models are explained in more details in sections 4.3 and 4.4.

Bounding box containing object of interest was specified for every image. This form of initialization is a more simple alternative to providing seeds for object and background.

4.1 Unary and pairwise terms

In our experiments color models for object and background were represented by mixtures of Gaussians. We used approach proposed in [7] to learn color models using a bounding box of the object. The number of components in the mixture was set to 3.

For pairwise terms we used 4-connected neighborhood model $\mathcal{N}$. Terms were calculated as

$$\phi_{ij}(L_i, L_j) = \exp\left[-\lambda I[L_i \neq L_j]\left(e^{-c\frac{(B_i - B_j)^2}{\sigma^2}} + d\right)\right], \qquad (11)$$

where B_k represents color intensity for pixel k. Constant c was set to 1.2, d was set to 0.1, λ was set to 10 and σ was set to average difference between the intensities of neighboring pixels.

We used $\phi_i(L_i, S)$ of form (9) with $W(e, i)$ as in (10). Constants were set as $w = \ln 2$, $\alpha = 0.7$, $p = 2$.

4.2 Coordinate descent

As it was mentioned in section 3.3, S update step (4) was performed by simulated annealing. On the first update of S we initialized SA solution by automatically
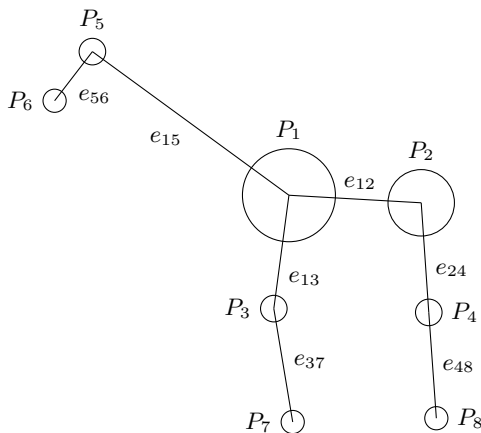
Fig. 5. Graph-based model of giraffe shape. Here $P_i = (P_i^x, P_i^y, P_i^r)$.

fitting most probable shape (the one that minimizes $E(S)$) into provided bounding box. Shape found on previous iteration was used as SA initialization for all the following S update steps.

Label update step (5) was performed via graph cuts. Function optimized during this step had form

$$F(L, S) = -\sum_i \log h_i(I_i, L_i) - w_s \sum_i \log \phi_i(L_i, S) - \\ - \sum_{i,j \in \mathcal{N}} \log \phi_{ij}(L_i, L_j). \tag{12}$$

We found that segmentation can be made more robust by smoothly increasing parameter w_s from 0 to 1 during first several iterations of coordinate descent. It can be viewed as a local minima avoidance heuristic.

We used labeling computed without shape prior ($w_s = 0$) as initial value for L. Coordinate descent stopped when the rate of pixels whose labels have changed after L update step was less than 0.0002. Shape influence w_s was linearly increased during first 10 iterations. Optimization process usually converged in $12 - 15$ iterations in our experiments and took about 2-3 minutes on a modern computer for a 320×240 image.

4.3   Giraffe segmentation

We applied our algorithm to a set of giraffe photos described above. The results of our algorithm were compared to segmentation without shape prior (initial labeling for our approach) and also with segmentation received by method from [7] that enforces tightness constraint on the segmented object.

Graph-based model of giraffe shape is presented in figure 5. We used the following expression for shape energy:

$$E(S) = \sum_{i=1}^{8} R_i \left(P_i^r, \sqrt{(P_1^x - P_2^x)^2 + (P_1^y - P_2^y)^2} \right) +$$
$$+E_{13}(e_{13}, e_{12}) + E_{24}(e_{24}, e_{12}) + E_{37}(e_{37}, e_{13}) +$$
$$+E_{48}(e_{48}, e_{24}) + E_{15}(e_{15}, e_{12}) + E_{56}(e_{56}, e_{15}). \tag{13}$$

In this expression term R_i constrains radius of the i-th vertex according to the length of giraffe body:

$$R_i(r, l; \rho_i, \sigma_i^r) = \frac{1}{\sigma_i^r}(r - \rho_i l)^2. \tag{14}$$

Here ρ_i specifies how radius of a particular vertex relates to the length of giraffe body and σ_i^r allows to control constraint softness. Pairwise terms $E_{ij}(e_1, e_2)$ establish constraints on relative length and angles between neighboring edges:

$$E_{ij}(e_1, e_2; \alpha_{ij}, \sigma_{ij}^\alpha, \rho_{ij}, \sigma_{ij}^l) =$$
$$= \frac{1}{\sigma_{ij}^\alpha}(\angle(e_1, e_2) - \alpha_{ij})^2 + \frac{1}{\sigma_{ij}^l}(\|e_1\| - \rho_{ij}\|e_2\|)^2. \tag{15}$$

Parameter α_{ij} specifies mean angle between edges, ρ_{ij} relates length of one edge to the length of another, parameters σ_{ij}^α and σ_{ij}^l control softness of the corresponding constraints. Energy function includes only relative constraints, and thus it is invariant to rotation and uniform scale of the shape. Values for all the parameters used in R_i and E_{ij} were selected manually.

Segmentation obtained by the proposed method for several giraffe photos is shown in figure 6. As we can see, in most cases initial segmentation includes many pixels with wrong labels. Graph-based shape prior seems to improve segmentation quality significantly. Tightness prior, on the other hand, is almost useless for pictures of this kind. Many segmentation errors occur near the boundaries of the bounding box and, therefore, make object even more tight.

Some typical situations when proposed method performs poorly are shown in figure 7. Left pair of images shows how bad initial segmentation can lead to coordinate descent solution that is far from desired optimum. Nevertheless, resulting segmentation is much closer to the ground truth than the initial one. Other images show situations when our hand-made shape model fails to handle all the shape variations, leading to a segmentation with some of the object pixels labeled as background. We think that more flexible shape model trained on labeled data can help to deal with such errors.

4.4   Letter segmentation

We also tested our algorithm on a number of images containing capital "E" letter. Shape model we used is shown in figure 8. Similar to giraffe shape model,
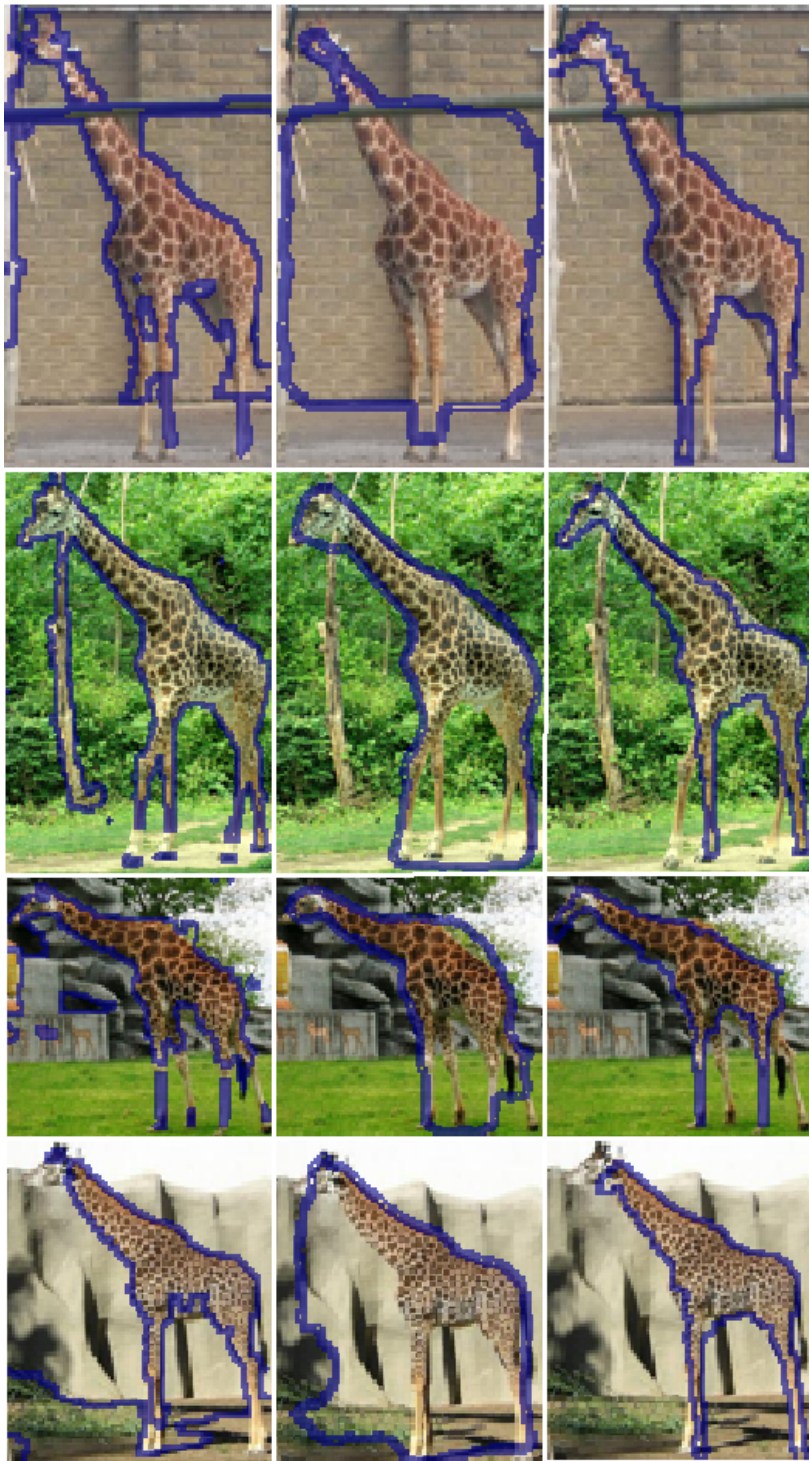
Fig. 6. Experimental results for giraffe images. Left: initial segmentation. Middle: segmentation with tightness prior. Right: segmentation with graph-based shape prior.
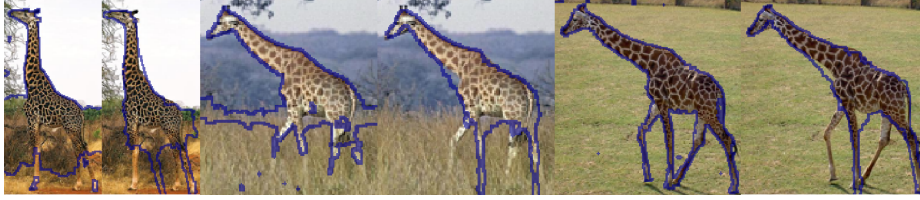
Fig. 7. Examples of bad segmentation with shape prior. Odd images: initial segmentation. Even images: segmentation with shape prior.

we used shape energy function of the form

$$E(S) = \sum_{i=1}^{5} R_i \left(P_i^r, \sqrt{(P_1^x - P_3^x)^2 + (P_1^y - P_3^y)^2} \right) + \tag{16}$$
$$+ E_{12}(e_{12}, e_{25}) + E_{23}(e_{23}, e_{25}) + E_{14}(e_{14}, e_{12}) + E_{36}(e_{36}, e_{23}),$$

with R_i defined in (14) and E_{ij} defined in (15). Results of segmenting letter images with and without shape prior are shown in figure 9. As with giraffe photos, shape prior has improved segmentation quality significantly.
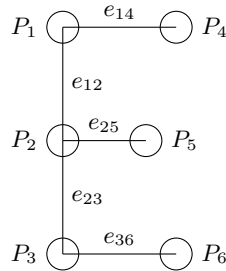
Fig. 8. Graph-based model of capital "E" letter

5   Conclusion

In this paper we present an iterative approach to image segmentation with a shape prior. Approach is based on a posterior probability maximization via coordinate descent and can be seen as a degenerate kind of EM algorithm. Each iteration of coordinate descent consists of two stages: shape fitting via simulated annealing and image segmentation with re-estimated unary terms.

We also propose a shape prior that is applicable to objects with well-defined structure. Presented shape model is a graph with variable width specified for every edge. Such representation allows to control shape variation and to specify

Fig. 9. Algorithm results for letter images. Left: segmentation without shape prior. Middle: segmentation with shape prior. Right: shape skeleton obtained during segmentation.

global constraints like known orientation or scale of an object. We show how one can build such a prior into proposed segmentation scheme. Experiments confirm that proposed shape prior can make segmentation less sensitive to the lack of reliable information about object edges and color.

Finding more efficient shape matching technique for S update step that would replace simulated annealing can become one direction of future research. Possible options include DP-based approach similar to the one used in [12] and branch-and-bound technique [4]. It is, however, questionable if branch-and-bound can help to improve processing speed significantly. This question requires further investigation.

It would also be useful to exclude the manual stage of shape model creation. One can try to learn the structure of the shape graph together with the parameters controlling its flexibility from a set of manually segmented images.

References

1. Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In: 2001 IEEE 8th International Conference on Computer Vision. Volume 1., IEEE (2001) 105–112
2. Vu, N., Manjunath, B.: Shape prior segmentation of multiple objects with graph cuts. 2008 IEEE Conference on Computer Vision and Pattern Recognition (2008) 1–8
3. Freedman, D., Zhang, T.: Interactive Graph Cut Based Segmentation with Shape Priors. 2005 IEEE Conference on Computer Vision and Pattern Recognition (2005) 755–762
4. Lempitsky, V., Blake, A., Rother, C.: Image segmentation by branch-and-mincut. Proceedings of the 10th European Conference on Computer Vision (2008) 15–29
5. Cremers, D., Schmidt, F.R., Barthel, F.: Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. (2008)
6. Veksler, O.: Star Shape Prior for Graph-Cut Image Segmentation. In: Proceedings of the 10th European Conference on Computer Vision. (2008)
7. Lempitsky, V., Kohli, P., Rother, C., Sharp, T.: Image segmentation with a bounding box prior. In: Computer Vision, 2009 IEEE 12th International Conference on, IEEE (2009) 277–284
8. Kumar, M.P., Torr, P.H., Zisserman, A.: Obj Cut. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Volume 1., IEEE Computer Society (2005)
9. Bray, M., Kohli, P., Torr, P.H.: Posecut: Simultaneous segmentation and 3d pose estimation of humans using dynamic graph-cuts. Proceedings of the 8th European Conference on Computer Vision **01** (2006) 642–655
10. Latecki, L.J.: Active skeleton for non-rigid object detection. In: 2009 IEEE 12th International Conference on Computer Vision, IEEE (2009) 575–582
11. Quack, T., Ferrari, V., Leibe, B., Van Gool, L.: Efficient Mining of Frequent and Distinctive Feature Configurations. 2007 IEEE 11th International Conference on Computer Vision (2007)
12. Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial Structures for Object Recognition. International Journal of Computer Vision **61** (2005) 55–79