Combining Shape Priors and MRF-Segmentation

Boris Flach and Dmitrij Schlesinger

Dresden University of Technology*

Abstract. We propose a combination of shape prior models with Markov Random Fields. The model allows to integrate multiple shape priors and appearance models into MRF-models for segmentation. We discuss a recognition task and introduce a general learning scheme. Both tasks are solved in the scope of the model and verified experimentally.

1 Introduction

The last decade's progress in image analysis and image interpretation is strongly influenced by a model driven attitude. In particular this applies for image segmentation, whereas the goal is to partition the image domain into meaningful areas.

In particular discrete structural models and approaches were strongly influenced by the breakthrough achieved in the area of $(\max, +)$-Labelling problems also known as Soft Constraint Satisfaction problems. Findings on polynomially solvable classes of these problems [1,2,3] and deduction of well-founded approximation algorithms for rather general classes [4,5,6] have led to considerable progress for image segmentation (and other areas of image analysis as well).

The adaption of probabilistic shape priors from level-set based models [7,8] and their further development and combination with simple segmentation models [9] and later on with Markov Random Fields and Conditional Random Fields [10,11] marks another leap in the segmentation research.

A closer examination of particular aspects like model complexity, recognition and learning reveals however, that many of the above mentioned approaches concentrate on some of these aspects and leave out others at the same time. Some of these approaches combine shape prior models with simple segmentation models – i.e. without lateral interactions – and solve the corresponding segmentation and learning tasks [9]. Other approaches combine very complex shape models and CRF-s and solve corresponding recognition tasks as well as supervised learning. Partially unsupervised learning for CRF-s (without shape priors) is considered in [12]. However, CRF-s have some principal drawbacks. They cannot be learned completely unsupervised. The same holds, if the observation is partially hidden. Moreover, it is yet unclear how to measure their generalisation capability (e.g. VC-Dimension). Finally, the majority of papers propose the MAP-criterion for segmentation, which is not natural – using the Hamming distance for loss is much more appropriate.

* Supported by Deutsche Forschungsgemeinschaft, Grant No. FL307/2-1.

N. da Vitora Lobo et al. (Eds.): SSPR&SPR 2008, LNCS 5342, pp. 177–186, 2008.

In the present paper we propose a probabilistic segmentation model based on Markov Random Fields and object/segment specific shape priors. The latter range midscale in complexity. The model combines the previously mentioned benefits and avoids unnecessary restrictions at the same time. It allows to pose *all* variants of recognition as well as *all* variants of learning in a closed fashion and to solve them approximatively.

2 The Model

For the sake of convenience we consider throughout this paper a relatively simple model for the shape priors – an expected average shape defined up to pose parameters like position and scale. Denoting the discrete image domain by $R \subset \mathbb{Z}^2$, the sample space Ω of the probabilistic model is $\mathcal{X} \times \mathcal{S}$, where $\mathcal{X}$ is the set of all images

$$x \colon R \to F, \quad F \text{ - colour space} \tag{1}$$

and $\mathcal{S}$ is the set of all segmentations into K regions

$$s \colon R \to K, \quad K \text{ - label/segment set.} \tag{2}$$

A shape prior function $\phi_k(r, \mu)$ is assigned to each object/segment, where

$$\phi_k(., \mu) \colon \mathbb{R}^2 \to \mathbb{R} \tag{3}$$

is the signed distance to the average object contour defined up to a set of pose parameters μ. The better a segmentation $s \in \mathcal{S}$ fits the prior shape, the more probable it is. The same is required with respect to "compactness". Modelling the latter by Potts-interactions, gives the following prior model for segmentations

$$P(s) = \frac{1}{Z} \exp\Big[\alpha \sum_{\{rr'\} \in E} \delta(s_r, s_{r'}) + \lambda \sum_{r \in R} \sum_{k \in K} \delta(s_r, k) \phi_k(r, \mu_k)\Big] , \tag{4}$$

where $\delta(.,.)$ denotes the Kronecker symbol and E is a neighbourhood relation on R. Z denotes the partition function and depends on α, λ, $\bar{\mu} = \{\mu_k, k \in K\}$ and the choice of $\phi_k(.,.)$.

The second part of our probabilistic model – the conditional probability to observe a certain image given a segmentation – is assumed to be pixelwise independent:

$$P(x \mid s) = \prod_{r \in R} q(x_r \mid s_r) , \tag{5}$$

where $q(f \mid k)$, $f \in F$, $k \in K$ are arbitrary conditional p.d-s describing the appearance of the object segments.

Let us remark here, that the idea to adopt signed distance functions (from level set approaches) for discrete probabilistic models is not new and was proposed e.g. in [9]. Our model is nevertheless much more complex due to the presence of Potts-interactions. On the other hand, even more complex models where

proposed and used in the context of Conditional Random Fields (see e.g. [10]). Unfortunately, the good recognition results achievable by CRF-s have their price – the latter have drawbacks e.g. with respect to learning. The proposed model, ranging midscale with respect to complexity, is *generative* and allows to pose arbitrary learning and recognition tasks.

3 Recognition

In this section we study the task of pose and segmentation estimation for a given image $x \in \mathcal{X}$. We begin with pose estimation. We consider it as an unknown model parameter and estimate it by the Maximum Likelihood principle[1]

$$\bar{\mu}_* = \arg\max_{\bar{\mu}} P(x;\bar{\mu}) = \arg\max_{\bar{\mu}} \sum_{s\in\mathcal{S}} P(x,s;\bar{\mu})\,. \tag{6}$$

At this point we would like to explain, why we advocate this approach instead of the usually used

$$\bar{\mu}_*, s_* = \arg\max_{\bar{\mu},s} P(x,s;\bar{\mu})\,. \tag{7}$$

We assume for simplicity, that we are interested in the pose parameters only; i.e. the segmentation does not matter. The latter is not known, since only the image x is observed. Hence, it is reasonable to sum over all segmentations, weighting them by their probabilities $P(x,s;\bar{\mu})$, in order to evaluate a certain choice of $\bar{\mu}$.

Currently there are no efficient algorithms for the calculation of the required sum over the vast number of segmentations in (6). It is reasonable to use the EM-scheme here, what gives the iteration

$$\begin{aligned} &\text{E:}\quad \beta^{(n)}(s) = P\big(s \mid x;\bar{\mu}^{(n)}\big) \\ &\text{M:}\quad \bar{\mu}^{(n+1)} = \arg\max_{\bar{\mu}} \sum_{s\in\mathcal{S}} \beta^{(n)}(s)\log P(x,s;\bar{\mu})\,. \end{aligned} \tag{8}$$

Substitution of the used model in M-step gives

$$L = \lambda \sum_{r\in R}\sum_{k\in K} \beta^{(n)}(s_r=k)\phi_k(r,\mu_k) - \log Z(\bar{\mu}) \to \max_{\bar{\mu}}\,. \tag{9}$$

Currently it is impossible to perform this maximisation directly, because it is unknown how to calculate the partition sum. Interestingly, it is possible to derive its gradient with respect to $\bar{\mu}$, what gives:

$$\frac{\partial L}{\partial \mu_k} = \lambda \sum_{r\in R}\Big[P\big(s_r=k \mid x;\bar{\mu}^{(n)}\big) - P\big(s_r=k;\bar{\mu}\big)\Big]\,\frac{\partial \phi_k(r,\mu_k)}{\partial \mu_k}\,. \tag{10}$$

Thus it is necessary to calculate the difference of marginal a-posteriori and a-priori probabilities for the segment label $k \in K$ in every pixel $r \in R$ in order to perform the gradient step.

[1] Hereby and in what follows, we adopt the convention that parameters of a p.d. are separated from events by a semicolon.

Let us now return to the question how to estimate an optimal segmentation. We formulate this question as a task of Bayesian decision. Thereby we adopt the Hamming distance between the true and the estimated segmentation for the loss function. The resulting optimal Bayes decision strategy is

$$s_r^* = \arg\max_{k \in K} P(s_r = k \mid x; \bar{\mu}) \,. \tag{11}$$

Again, all we need for the estimation of the optimal segmentation are the posterior marginal probabilities for the segment labels.

It is of course clear that the attainable precision of the whole recognition scheme depends on the remaining two model parameters α and λ. Moreover, we expect a certain interplay of these parameters because the shape priors ϕ_k are assumed to be signed distance functions for the expected contours (or a sign preserving monotone function of the latter). In order to study this effect and to evaluate the learning results given in the next section, we conducted the following series of experiments. Assuming a circular shaped object

$$\phi(r) = \frac{\rho^2 - \|r - z\|^2}{2\rho} \tag{12}$$

where ρ is the radius of the circle and z is its centre position, seven images were created as follows. The contour of the circle was deformed and the resulting two-valued images were distorted by replacing the greyvalue in 80% of randomly chosen pixels by uniformly distributed noise. Figure 1 shows two of these pairs. We applied the whole recognition scheme including the estimation of the pose and segmentation to all seven images for different combinations of λ and α. The initial poses were chosen so as to have a substantial overlap of the circles – the centres were initialised randomly inside the true circle. For each α, λ combination and each image we get the estimated pose parameters as a result of the EM-algorithm. At the same time the segmentation is estimated according to (11). In addition, MAP-segmentation was determined after pose convergence for comparison. Except the case of $\lambda = 0$ the pose parameters were estimated correctly. Table 1 shows the averaged Hamming distance between the ground truth and the segmentation obtained by (11) and by MAP-decision (upper and lower number in each cell respectively). First off all, it is seen, that for $\alpha = 0$ (i.e. absence of the Potts-prior), the result is the better, the "stronger" the prior shape

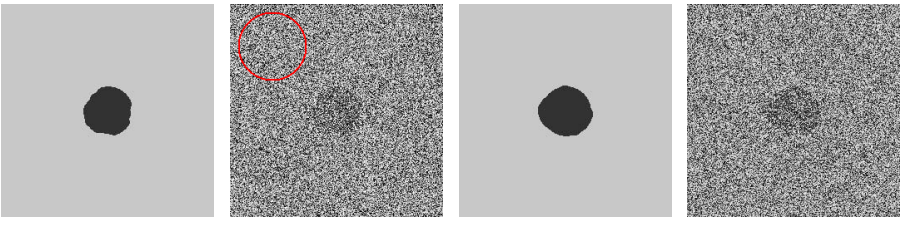

Fig. 1. Two of the images pairs used for experiments. Each is shown with ground truth on the left and the distorted image on the right. (See Section 5 for the meaning of the red circle).

Table 1. Average Hamming distance between the estimated segmentations and ground truth

$\lambda\backslash\alpha$	0	0.7	1.4	1.8	2.1	2.3	3.0
0.00	1.e4 1.e4	1015 360	107 162	81 155	82 155	86 162	92 158
0.03	1405 1366	189 186	**85** 124	83 129	**82** 139	92 143	95 159
0.06	1075 1081	151 147	86 110	**80** 120	87 122	**83** 125	92 136
0.10	647 627	126 132	93 110	86 114	90 112	84 112	**89** 121
0.30	214 206	**125** 126	113 120	101 113	104 115	98 113	103 110
0.60	187 186	139 142	130 132	123 128	126 135	129 132	136 139
1.00	**183** 183	152 154	146 148	145 144	148 148	147 153	148 152

model i.e. the bigger the λ is chosen. This behaviour changes for $\alpha > 0$ – for each such α there is an optimal λ-interval inside the analysed range. The overall optimum is reached at $\alpha > 0$, what demonstrates the usefulness of the Potts-prior. And finally, the Bayes decision (11) clearly outperformes the MAP decision.

4 Learning

This section is devoted to learning. Given a learning sample, the task is to estimate the remaining model parameters α and λ and/or the appearance characteristics $\bar{q}$. Especially in the context of structural pattern recognition it is highly desirable to utilise for learning not only completely segmented examples but also partially segmented examples as well as example images without segmentation and even partially observed images with or without (partial) segmentation. Learning of model parameters from partially segmented images (without shape priors) was considered in [12] in the context of CRF-s. However, in order to recognise or learn, CRF-s require *complete* observations and at least partial segmentations for learning, whereas generative models like ours can cope with partial observations as well as with completely absent segmentations.

We propose to consider a learning sample T as a set of *events* $A_j \subset \Omega = \mathcal{X} \times \mathcal{S}$, $j = 1, 2 \ldots, \ell$. The learning task, if considered according to the Maximum Likelihood Principle, reads then

$$L = \log P(T; \alpha, \lambda, \bar{q}) = \sum_{j=1}^{\ell} \log \sum_{(x,s)\in A_j} P(x, s; \alpha, \lambda, \bar{q}) \to \max_{\alpha,\lambda,\bar{q}} . \qquad (13)$$

Clearly, all previously mentioned situations are absorbed in this formulation.[2]

[2] Please note, that an event A_j might typically consist of a huge number of elementary events.

Because the log-likelihood in (13) cannot be calculated and maximised directly, we again use the EM-scheme here. We introduce non negative coefficients

$$\beta_j(x,s) \geqslant 0 \;\; \forall (x,s) \in A_j \quad \text{where} \sum_{(x,s)\in A_j} \beta_j(x,s) = 1 \tag{14}$$

for each example and write the log-likelihood L as

$$L(m) = \sum_{j=1}^{\ell} \sum_{(x,s)\in A_j} \beta_j(x,s) \log P(x,s;m) \\ - \sum_{j=1}^{\ell} \sum_{(x,s)\in A_j} \beta_j(x,s) \log P(x,s \mid A_j;m) \tag{15}$$

where m subsumes all model parameters $m = \{\alpha, \lambda, \bar{q}\}$ or a subset of them. Choosing now the coefficients β according to

$$\beta_j^{(n)}(x,s) = P(x,s \mid A_j; m^{(n)}) \tag{16}$$

in each E-step of the EM-iteration, ensures that the subtrahend in (15) has a global maximiser at $m = m^{(n)}$. Maximising (or at least increasing) the minuend, ensures an increase of the log-likelihood. The task to be solved in the M-step is thus to maximise the minuend $\sum_j L_j$ in (15) with respect to m. Substituting our model (4), (5) and denoting

$$\begin{aligned} \beta_j^{(n)}(s_r = s_{r'}) &= P(s_r = s_{r'} \mid A_j; \alpha^{(n)}, \lambda^{(n)}, \bar{q}^{(n)}) \\ \beta_j^{(n)}(s_r = k) &= P(s_r = k \mid A_j; \alpha^{(n)}, \lambda^{(n)}, \bar{q}^{(n)}) \,. \end{aligned} \tag{17}$$

gives the following results. The new appearance characteristics are (up to a straightforward normalisation)

$$q^{(n+1)}(f \mid k) \sim \sum_j \sum_{r \in R_j(f)} \beta_j^{(n)}(s_r = k) \,, \tag{18}$$

where $R_j(f)$ denotes the set $\{r \in R \mid x_r^j = f\}$.

Similar as for recognition, it is impossible to maximise $\sum_j L_j$ directly, with respect to α and λ because it is unknown how to calculate the partition sum efficiently. Nevertheless, it is possible to derive its gradient, what gives

$$\begin{aligned} \frac{\partial L_j}{\partial \alpha} &= \sum_{\{rr'\}\in E} \Big[\beta_j^{(n)}(s_r = s_{r'}) - P(s_r = s_{r'}; \alpha, \lambda, \bar{q}) \Big] \\ \frac{\partial L_j}{\partial \lambda} &= \sum_{r\in R} \sum_{k\in K} \Big[\beta_j^{(n)}(s_r = k) - P(s_r = k; \alpha, \lambda, \bar{q}) \Big] \phi_k(r, \mu_k) \,. \end{aligned} \tag{19}$$

In order to perform the M-step in its gradient version, we need a-priori and a-posteriori marginal probabilities conditioned by the corresponding events A_j.

In order to verify the reliability of the derived learning scheme with respect to α and λ, we accomplished a series of experiments. The true deformed circular segmentations (see Sec. 3) were used as a learning sample. This corresponds to the supervised version of learning. Because each A_j consits of only one segmentation s in this case, the posterior marginal probabilities β in (17) are known in advance. Learning of λ was performed for certain fixed α-values. The results are shown in Table 2. Comparison with the recognition results given in Sec. 3 shows an almost perfect agreement.

Table 2. Values of λ learned for different α

α	0	0.7	1.4	1.8	2.1	2.3	3.0
λ	0.91	0.32	0.07	0.05	0.048	0.046	0.091

Completely unsupervised learning of the appearance characteristics $\bar{q}$ (on a per image basis) was used and thus verified in experiments with real images (see next section).

5 Other Experiments

We have used rather good pose initialisations in all experiments described so far. In order to study the ability to cope with wrong initial pose estimates, we made the following experiment. Using one of the previously described images, the recognition scheme for pose estimation and segmentation was combined with unsupervised learning of the appearance characteristics $\bar{q}$ and the shape parameter λ. The latter was fixed to a small value at the beginning and started to learn after some delay. The initial pose was located in the upper left corner (shown as the red circle in Fig. 1), thus having no overlap with the true segment. The results of the experiment are shown in Fig. 2. At the beginning of the experiment (λ is fixed to a small value) the appearance characteristics $\bar{q}$ are learned

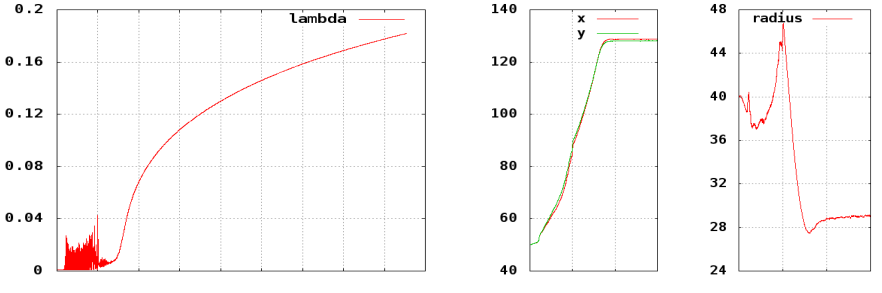

Fig. 2. Results of recognition and simultaneous learning of λ: learning curves for λ, the centre position and the radius respectively (the constant parts of the latter two were clipped)

Fig. 3. Original images, segmentations with overlaid shapes, segmentations without shape prior

approximately and the segmentation shows two blobs - located at true position and at the actually estimated (wrong) pose respectively. A rather interesting phenomenon is observed when λ starts to learn: The blobs "compete" with each other and λ oscillates. At the same time the pose starts to approach the true one ($x = 128$, $y = 128$ and $\rho = 30$). Finally, the pose is correctly estimated and λ approaches a reasonable value ($\alpha = 0.7$ was used in this experiment).

The next examples (Fig. 3) show segmentation results for real images. A rather good pose initialisation was used in all experiments. The appearance characteristics $\bar{q}$ were assumed to be a mixture of multivariate Gaussians and learned completely unsupervised. The number of segments and the numbers of Gaussians for each segment (typically from 2 to 4) as well as the parameters α and λ were estimated experimentally. The appearance model includes additional a-priori slowly varying shading fields for each segment (see [13] for details). These shadings were learned simultaneously with $\bar{q}$ in a fully unsupervised manner. These experiments clearly demonstrate the usefullness of shape priors, especially for learning.

6 Conclusions and Open Questions

We introduced a combination of shape prior models with Markov Random Fields. The model allows to integrate multiple shape priors and appearance models into MRF-models for segmentation. At the same time it allows to pose various recognition tasks and a wide spectrum of learning tasks. Our main goal was to pay attention to concise and reasonable formulation of these tasks as opposed to increasing the model complexity. As demonstrated in experiments, segmentation of real images is greatly improved, even if relatively simple shape priors are used.

Yet there are several open questions which we are currently investigating or want to investigate in the future:

In the current model we considered the pose of a shape as a parameter. Moreover, only "templates" were used for the shape. It seems to be beneficial to extend these models towards statistical shape models (like e.g. in [9]). At the same time this would mean to consider other types of recognition tasks, based on Bayes decision for the pose and possibly for the shape.

It is yet unclear how to cope with wrong pose initialisations. The experiment presented in the paper is a first trial at most. In case of statistical shape models, the pose and the shape can be sampled (like e.g. in [10]), but it is still unclear, whether such a replacement of the gradient scheme will solve the problem efficiently.

References

1. Greig, D.M., Porteous, B.T., Scheult, A.H.: Exact maximum a posteriori estimation for binary images. Journal of the Royal Statistical Society B 51(2), 271–279 (1989)
2. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2352, pp. 65–81. Springer, Heidelberg (2002)

3. Schlesinger, D., Flach, B.: Transforming an arbitrary minsum problem into a binary one. Technical Report TUD-FI06-01, Dresden University of Technology (April 2006)
4. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. In: Proc. of the 7. Intl. Conf. on Computer Vision, vol. 1, pp. 377–384 (1999)
5. Kovtun, I.: Partial optimal labeling search for a NP-hard subclass of (max,+) problems. In: Michaelis, B., Krell, G. (eds.) DAGM 2003. LNCS, vol. 2781, pp. 402–409. Springer, Heidelberg (2003)
6. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(10), 1568–1583 (2006)
7. Leventon, M.E., Grimson, W.E., Faugeras, O.: Statistical shape influence in geodesic active contours. In: Proc. IEEE Conf. Comp. Vision and Patt. Recog. (2000)
8. Cremers, D., Sochen, N., Schnörr, C.: A multiphase dynamic labeling model for variational recognition-driven image segmentation. International Journal of Computer Vision 66(1), 67–81 (2006)
9. Pohl, K.M., et al.: Shape based segmentation of anatomical structures in magnetic resonance images. In: Liu, Y., Jiang, T., Zhang, C. (eds.) CVBIA 2005. LNCS, vol. 3765, pp. 489–498. Springer, Heidelberg (2005)
10. Kumar, M.P., Torr, P.H.S., Zisserman, A.: An object category specific MRF for segmentation. In: Ponce, J., Hebert, M., Schmid, C., Zisserman, A. (eds.) Toward Category-Level Object Recognition. LNCS, vol. 4170, pp. 596–616. Springer, Heidelberg (2006)
11. Levin, A., Weiss, Y.: Learning to combine bottom-up and top-down segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 581–594. Springer, Heidelberg (2006)
12. Verbeek, J., Triggs, B.: Scene segmentation with CRFs learned from partially labeled images. In: Advances in Neural Information Processing Systems, vol. 20 (2008)
13. Schlesinger, D., Flach, B.: A probabilistic segmentation scheme. In: Rigoll, G. (ed.) DAGM 2008. LNCS, vol. 5096, pp. 183–192. Springer, Heidelberg (2008)