A Probabilistic Segmentation Scheme

Dmitrij Schlesinger and Boris Flach

Dresden University of Technology

Abstract. We propose a probabilistic segmentation scheme, which is widely applicable to some extend. Besides the segmentation itself our model incorporates object specific shading. Dependent upon application, the latter is interpreted either as a perturbation or as meaningful object characteristic. We discuss the recognition task for segmentation, learning tasks for parameter estimation as well as different formulations of shading estimation tasks.

1 Introduction

Segmenting images in meaningful parts is a common task of image analysis. "Common" means that segmentation arises as a subtask in the context of diverse and different applications of image analysis and computer vision. Corresponding research efforts of the previous decades – often motivated by these different contexts – have led to a plenty of segmentation algorithms and their variants. Unfortunately, this does not hold for the number of corresponding models – the algorithms were often constructed in a rather phenomenological manner. This discrepancy is even greater if either supervised or unsupervised learning is required for the parameters of the algorithm/model.

On the other hand, by now there is no hope for a universal model/algorithm pair for segmentation. Since segmentation is often a subtask of a recognition task, it provides rather intermediate results, which are determined not by the image(s) only, but also e.g. by feedback from "higher" model parts (like object models). Hence, we believe that segmentation schemes, which are to some extend widely applicable, should have the following properties. Firstly, there should be a clear ab initio model (preferable a modular one). Secondly, it is necessary to have well-posed recognition and learning task formulations. Finally, the scheme should have interfaces (like model parameters) for feedback from higher model levels. These requirements suggest probabilistic models. The main reason for this choice is learning – statistical pattern recognition has the farthermost advanced theory of learning. Besides, it is often preferable to have a (posterior) probability distribution of segmentations instead of a unique segmentation.

Unfortunately, the choice of a probabilistic model presently disallows continuous variational approaches like e.g. Level Sets [3,2]. These models are not extensible to stochastic ones, mainly because it is not yet clear, how to define probability measures on function spaces correctly.

The actually most popular approach in the scope of discrete models for segmentation is Energy Minimisation [6,1,5,7]. In most cases it corresponds to the Maximum A-Posteriori Decision with respect to a certain probabilistic model. We believe however, that an overhastily preference for the MAP-criterion may hide other possible formulations of recognition tasks and – what is more important – reasonable approaches for learning.

In the following we present a scheme for segmentation. Its main part is a probabilistic model. In particular we use Gibbs probability distributions of second order to represent the distribution of hidden variables. The second part of the method is the formulation of recognition tasks. We would like to point out, that the latter can be formulated and derived in many different ways within the same model. Last but not least we consider the learning of unknown parameters of the probability distribution. We give an unsupervised learning scheme, which is based on the Maximum Likelihood principle. In particular we use the Expectation Maximisation algorithm to solve learning tasks approximatively.

2 The Model

The sample space of the model is built by triples of the following groups of variables: the image(s), the segmentation field and segment related shading fields. We decided to include the shading into the model due to the following reasons. Firstly, shading is often a segment specific concomitant phenomenon. Secondly, in some applications (e.g. Shape from Shading) it is rather a quantity we are interested in, as opposed to just a perturbation variable. Finally, for one and the same model, the task of shading estimation might differ, depending on the application.

Throughout the paper we use the following notation. The field of vision $R \subset \mathbb{Z}^2$ is equipped with the structure of an undirected graph $\mathcal{G} = (R, E)$, where E denotes the set of edges. S is a finite set of segment labels, where $s \in S$ denotes the label of a particular segment. The segmentation field $f : R \to S$, where $f(r) \in S$ denotes the segment chosen for the element $r \in R$ of the field of vision. An observation (i.e. image) is denoted by $x : R \to V$, where V is a set of colour values. And finally, $h_s : R \to V$ denotes the shading field associated with segment s. The set of all shadings is $h = (h_1, h_2 \dots h_{|S|})$. The probability of an elementary event – a triple (f, h, x) is modelled by

$$p(f,h,x) = p(f) \cdot p(x \mid f,h) \cdot \prod_{s} p(h_s).$$
(1)

The prior probability distribution for segmentations is assumed to be a Gibbs probability distribution of second order

$$p(f) \sim \prod_{rr' \in E} g_{segm} \left(f(r), f(r') \right)$$
(2)

with $g_{segm}: S \times S \to \mathbb{R}^+$. A particular and appropriate choice for segmentation problems is e.g. the Potts model

$$g_{segm}(s,s') = \begin{cases} a > 1 & \text{if } s = s' \\ 1 & \text{otherwise.} \end{cases}$$
(3)

Similarly, the prior probability of a shading field is

$$p(h_s) \sim \prod_{rr' \in E} g_{shad} \big(h_s(r), h_s(r') \big), \tag{4}$$

where $g_{shad}: V \times V \to \mathbb{R}^+$ expresses the prior assumptions about shadings, e.g. smoothness or hard restrictions:

$$g_{shad}(v,v') = \exp\left[-(v-v')^2/2\sigma^2\right],$$
 (5)

$$g_{shad}(v,v') = \mathbf{1}\{|v-v'| < \delta\}.$$
 (6)

The probability distribution of an observation, given the other two variables, is assumed to be conditionally independent

$$p(x \mid f, h) = \prod_{r} q(x(r) - h_{f(r)}(r), f(r)).$$
(7)

The expression $x(r) - h_{f(r)}(r)$ can be understood as "shading adjusted observation value" at position r and is obtained by subtracting the shading value associated with the segment f(r) chosen in this node. The function $q: V \times S \to \mathbb{R}^+$ is a conditional probability p(v|s) for the colour value v given the segment s.

3 Recognition and Learning Tasks

In this section we discuss different tasks, which can be formulated in the scope of the model. We begin with a typical recognition task – segmentation estimation. Afterwards, we give a scheme for unsupervised learning of the unknown conditional probability distribution q, given an example image. Finally, we consider the shading estimation. It turns out, that both "recognition" and "learning" formulations are reasonable for shading estimation. The choice depends on the particular application. We discuss both variants and compare them.

3.1 Recognition

Let us assume that the parameters of the model (e.g. functions g and q) are known. The shadings h_s are considered as a perturbation and assumed to be known as well. The task is to estimate the segmentation f^* given an observation x. We formulate the segmentation problem as a task of Bayesian decision, i.e. we minimise the risk

$$R(f) = \sum_{f'} p(f' \mid h, x) \cdot C(f, f') \to \min_{f},$$
(8)

where C(f, f') is the loss function. When applying decision theory for pattern classification, it is common to use a (possibly class dependent) constant for the loss associated to a classification error. The situation is quite different in the scope of segmentation: it is reasonable to use e.g. the number of mis-segmented

pixels for the loss function. (Note that this loss is usually used when comparing different segmentation algorithms.) Hence, we advocate an additive loss function of the type

$$C(f, f') = \sum_{r} c(f(r), f'(r)),$$
(9)

where $c: S \times S \to \mathbb{R}$ is a function, which penalises deviations from the estimated segment label in a node r to the unknown true one. Substituting a loss function of that type in (8) gives

$$R(f) = \sum_{f'} p(f' \mid h, x) \cdot \sum_{r} c(f(r), f'(r)) = \sum_{r} \sum_{f'} p(f' \mid h, x) \cdot c(f(r), f'(r)).$$

Because the second factor depends only on the segment label of the node r, it is possible to split the sum over all segmentation fields f' into the sum over all $f'(R \setminus r)$ and the sum over all segmentation labels of the node r. This gives

$$R(f) = \sum_{r} \sum_{s \in S} c(f(r), s) \cdot p(f(r) = s \mid h, x) \to \min_{f},$$
(10)

where

$$p(f(r) = s \mid h, x) = \sum_{f': f'(r) = s} p(f' \mid h, x)$$
(11)

are the a-posteriori marginal probability distributions of states. The optimisation (10) can be performed for each node r independently and gives

$$f^*(r) = \underset{s}{\operatorname{arg\,min}} \sum_{s'} c(s,s') \cdot p\big(f(r) = s' \mid h, x\big) \text{ for each } r.$$
(12)

In particular the additive delta cost function – i.e. $c(s, s') = \mathbf{1}\{s \neq s'\}$ in (9) – can be used for segmentation. This leads to the decision

$$f^*(r) = \operatorname*{arg\,max}_{s} p\big(f(r) = s \mid h, x\big) \quad \text{for each} \quad r.$$
(13)

It is easy to see, how this decision should be changed if the shading fields are unknown:

$$f^*(r) = \arg\max_{s} p(f(r) = s \mid x) = \arg\max_{s} \sum_{h} \sum_{f': f'(r) = s} p(f', h \mid x).$$
(14)

3.2 Learning

Let us consider the task of unsupervised learning of unknown parameters of the probability distribution – e.g. learning the conditional probability distributions q given an image. For simplicity, we still assume that the shadings h_s are known. We follow the Maximum Likelihood principle and maximise¹

$$\ln p(x \mid h; q) = \ln \sum_{f} p(f, x \mid h; q) \to \max_{q}.$$
(15)

¹ The notation $p(\ldots;q)$ means "parameterised by q".

We use the EM-algorithm for approximation. The standard approach leads to the following iterative scheme:

- 1. Expectation step: compute the a-posteriori probability distribution of segmentations for the current set of parameters, i.e. $p(f|h, x; q^{(n)})$;
- 2. Maximisation step: maximise

$$q^{(n+1)} = \arg\max_{q} \sum_{f} p(f \mid h, x; q^{(n)}) \cdot \ln p(f, h, x; q).$$
(16)

Substitution of the last term in (16) by means of (1)-(7) and omission all terms, which do not depend on q, lead to

$$\sum_{f} p(f \mid h, x; q^{(n)}) \cdot \sum_{r} \ln q(x(r) - h_{f(r)}(r), f(r)) =$$

$$\sum_{s} \sum_{r} \sum_{f:f(r)=s} p(f \mid h, x; q^{(n)}) \cdot \ln q(x(r) - h_{s}(r), s) =$$

$$\sum_{s} \sum_{r} p(f(r) = s \mid h, x; q^{(n)}) \cdot \ln q(x(r) - h_{s}(r), s) \to \max_{q}.$$
(17)

Obviously, this optimisation can be performed for each segment independently. For a particular segment we have

$$\sum_{v} p(f(r) = s \mid h, x; q^{(n)}) \cdot \ln q(x(r) - h_s(r), s) =$$

$$\sum_{v} \left[\ln q(v, s) \cdot \sum_{r:x(r) - h_s(r) = v} p(f(r) = s \mid h, x; q^{(n)}) \right] \to \max_{q} .$$
(18)

Due to the Shannon's theorem, the solution of this task is

$$q^{(n+1)}(v,s) \sim \sum_{r:x(r)-h_s(r)=v} p(f(r)=s \mid h, x; q^{(n)}).$$
(19)

Summarising in a nutshell, an iteration of the learning algorithm is:

- 1. Expectation step: calculate the marginal a-posteriori probabilities of states in each node $p(f(r) = s|h, x; q^{(n)});$
- 2. Maximisation step:
 - (a) Sum up these marginals over all pixels with colour $x(r) = v + h_s(r)$ for each segment s and each colour value v, i.e. compute the "histogram"

$$hist(v,s) = \sum_{r:x(r)-h_s(r)=v} p(f(r) = s \mid h, x; q^{(n)}).$$
(20)

(b) Finally, normalise it:

$$q^{(n+1)}(v,s) = \frac{hist(v,s)}{\sum_{v'} hist(v',s)}.$$
(21)

Let us consider now the learning task for the case that the shadings are not known. In that case we have to optimise

$$\ln p(x;q) = \ln \sum_{f} \sum_{h} p(f,h,x;q) \to \max_{q}.$$
(22)

Although this task seems to be much more difficult than the previous one (15), it is not hard to see, that all derivations can be performed in a similar way. The difference is only, that it is necessary to sum over all possible shadings instead of considering them fixed. Finally, in the maximisation step of the EM-algorithm (19) the sum over *certain* pixels should be replaced by the sum over *all* pixels, weighted by corresponding marginal probabilities:

$$q^{(n+1)}(v,s) \sim \sum_{r} p(f(r) = s, x(r) - h_s(r) = v \mid x; q^{(n)}).$$
(23)

3.3 Shading Estimation

In contrast to the segmentation and learning tasks discussed so far, the situation with the shading is not so straightforward. In many applications the shading is considered as an unknown parameter of the model (caused e.g. by inhomogeneous and anisotropic lighting). This perturbation of object's appearance should be simply removed from the processed image. In such cases, it is reasonable to estimate the shading according e.g. to the Maximum Likelihood principle. Unfortunately, such a concept disallows incorporation of a-priori assumptions about shading in a "weighted" manner (like e.g. in (5)). The only possibility is to restrict the set of all possible shadings, using for example hard constraints like (6). Another way to deal with shading (which is in fact very similar to the first one), is to consider it as a statistical variable and to use the MAP criterion for its estimation.

In some applications it is however not reasonable to consider the shading as a perturbation because it may characterise certain properties of the segmented objects. Therefore, shading is not an "auxiliary" variable anymore. In such cases, it is appropriate to pose the problem of shading estimation again as a task of Bayesian decision. Hence, a suitable loss function should be chosen for that.

In the following we discuss both variants – MAP decision with respect to shading and shading estimation as a task of Bayesian decision. We begin with the first one. The task is to find

$$\ln p(h \mid x) = \ln \sum_{f} p(f, h \mid x) \to \max_{h}.$$
(24)

Again, we use the Expectation Maximisation algorithm to avoid summation over all segmentations. The standard approach leads to the following optimisation problem, which should be solved in each maximisation step:

$$\sum_{f} p(f \mid h^{(n)}, x) \cdot \ln p(f, h, x) \to \max_{h}.$$
(25)

We substitute our model for p(f, h, x) and obtain

$$\sum_{f} p(f \mid h^{(n)}, x) \cdot \left[\sum_{s} \ln p(h_{s}) + \sum_{r} \ln q(x(r) - h_{f(r)}(r), f(r)) \right] = \sum_{s} \left[\ln p(h_{s}) + \sum_{f} p(f \mid h^{(n)}, x) \sum_{r:f(r)=s} \ln q(x(r) - h_{s}(r), s) \right] \to \max_{h}.$$
 (26)

Obviously, the optimisation can be performed for each segment s separately:

$$\ln p(h_s) + \sum_{r} \sum_{f:f(r)=s} p(f \mid h^{(n)}, x) \cdot \ln q(x(r) - h_s(r), s) = \\ \ln p(h_s) + \sum_{r} p(f(r) = s \mid h^{(n)}, x) \cdot \ln q(x(r) - h_s(r), s) \to \max_{h_s}.$$
 (27)

Let us denote:

$$\tilde{g}(v,v') = -\ln g_{shad}(v,v') \text{ and}
\tilde{q}_r(v) = -p(f(r) = s \mid h^{(n)}, x) \cdot \ln q(x(r) - v, s).$$
(28)

The optimisation problem (27) can be written then in the form

$$\sum_{rr'\in E} \tilde{g}\big(h_s(r), h_s(r')\big) + \sum_r \tilde{q}_r\big(h_s(r)\big) \to \min_{h_s},\tag{29}$$

which is an Energy Minimisation task. If the function $\tilde{g}(.,.)$ is submodular – which is the case for (5) and (6) – then its solution can be found in polynomial time (see e.g. [8] for details).

Now let us discuss the variant that the shading is considered as a stochastic variable and should be estimated by defining an appropriate task of Bayesian decision

$$R(h) = \sum_{h'} p(h' \mid x) \cdot C(h, h') = \sum_{h'} \sum_{f} p(f, h' \mid x) \cdot C(h, h') \to \min_{h}.$$
 (30)

Again, we advocate an additive loss function of the type

$$C(h,h') = \sum_{s} \sum_{r} c(h_s(r), h'_s(r)).$$
(31)

Moreover, in the case of shading estimation the summands c(v, v') can be defined e.g. by $c(v, v') = (v - v')^2$, i.e. penalising deviations between the true and the estimated shading values. We omit here the derivation details for the decision strategy because of their similarity to the segmentation case. The Bayesian decision for the shading is its posterior mean value

$$h_s^*(r) = \sum_v v \cdot p(h_s(r) = v \mid x), \tag{32}$$

with the the marginal a-posteriori probabilities for shading values in each node

$$p(h_s(r) = v \mid x) = \sum_{h:h_s(r) = v} \sum_f p(f, h \mid x).$$
(33)

Remark 1. It is well known, that the calculation of marginal probabilities for Gibbs/Markov distributions is an NP-complete task. We need these probabilities for segmentation (12), learning (19) and shading estimation (28),(32). This illustrates once more, that the search for polynomial solvable subclasses of this problem is an important open question of structural pattern recognition. Until then, we are constrained to use approximative algorithms like the belief propagation method and others [9,10] or the Gibbs sampler [4].

4 Results

To begin with, we present an artificial example. The original image shown in Fig. 1a was produced by filling two segments with gradients, which have the same characteristics and finally adding Gaussian noise. The hard restrictions (6) were used for shading and the task of shading estimation was posed as Bayes decision task (30) with the loss function (31). The probability distributions q for segments were supposed to be zero mean Gaussians. Their variances were estimated unsupervised. The obtained segmentation is shown in Fig. 1c. The image in Fig. 1b was produced by replacing the original grayvalue in each pixel by the value of the estimated shading.

The next example (Fig. 2) is an image of a real scene. It is divided into three segments. The conditional probabilities q for the segments were supposed to be arbitrary, but channelwise conditionally independent. They were learned unsupervised by (23). The shading was handled in the same way as in the previous example. The image in Fig. 2c shows the result obtained for the same model but without shading fields.

The last example is related to satellite-based glacier monitoring and in particular to the recognition of debris covered glaciers of the everest-type. We applied our scheme to a visual and thermal infrared channel of the ASTER satellite



Fig. 1. An artificial example



(a) Original image(b) Segmentation(c) Without shadingsFig. 2. An example of a real scene



(a) Satellite image (b) Expert's segmentation (c) Segmentation

Fig. 3. Segmenting debris covered glaciers

combined with 3D elevation data. The conditional probabilities q were modeled by multivariate normal distributions for six segments and were learned partially unsupervised – during learning we used the expert's segmentation to permit three labels in the background, another two in the foreground (glaciers) and a last one in lake regions. Because of the homogeneous lighting, only one common vector-valued shading (modeled by (5)) was used for all segments and estimated by the MAP-criterion. Fig. 3 shows the visual satellite channel, the expert's segmentation (glaciers and lakes) and the obtained segmentation, which is correct for 96 percents of pixels.

5 Conclusion

We have presented a probabilistic scheme for segmentation which includes segment specific shading and admits well posed recognition and learning tasks.

In order to use the scheme for a particular application, it is necessary to decide, whether certain "parts" of the model should be considered as an (unknown) parameter or as a statistical variable. It is e.g. natural to consider the segmentation field as a statistical variable – because a particular segmentation is rather an event of a (possibly unknown) probability distribution and not a parameter which characterises a class of images.

The situation is different for the shading – both variants are suitable depending on the application. On the other hand, the probability distribution q was considered as a parameter throughout the paper. Though in principle possible, it is rather unnatural to consider it as a (unknown) statistical variable. The reason is twofold: often it is not possible to introduce a-priori assumptions for qand secondly, the choice for this function often corresponds to a class of images. Summarising, we suggest to consider a "part" of the model as a parameter either if it is hardly possible to formalise a-priori assumptions for it or if a particular instance of this part describes a class of images, and not a particular image.

For learning we have used the Maximum-Likelihood Principle. However, according to the learning theory other choices are possible. It is therefore highly desirable to analyse whether e.g. Minimisation of Empirical Risk can be generalised for structural pettern recognition and in particular for Gibbs/Markov probability distributions.

References

- 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)
- Cremers, D.: A multiphase level set framework for variational motion segmentation. In: Griffin, L.D., Lillholm, M. (eds.) Scale-Space Methods in Computer Vision (Isle of Skye), vol. 2695, pp. 599–614. Springer, Heidelberg (2003)
- Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. International Journal of Computer Vision 72(2), 195–215 (2007)
- Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6(6), 721–741 (1984)
- Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 28(10), 1568–1583 (2006)
- 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. 2350, pp. 65–81. Springer, Heidelberg (2002)
- 7. Micusik, B., Pajdla, T.: Multi-label image segmentation via max-sum solver. In: Proc. of Computer Vision and Pattern Recognition (2007)
- Schlesinger, D., Flach, B.: Transforming an arbitrary minsum problem into a binary one, Tech. report, Dresden University of Technology, TUD-FI06-01 (April 2005), http://www.bv.inf.tu-dresden.de/~ds24/tr_kto2.pdf
- Wainwright, M.J., Jaakkola, T.S., Willsky, A.S.: Tree-reweighted belief propagation algorithms and approximate ML estimation via pseudo-moment matching. In: Workshop on Artificial Intelligence and Statistics (January 2003)
- Wainwright, M.J., Jaakkola, T.S., Willsky, A.S.: A new class of upper bounds on the log partition function. IEEE Transactions on Information Theory 51(7), 2313– 2335 (2005)