Multiple Image Segmentation

Jonathan Smets and Manfred Jaeger^(✉)

Department for Computer Science, Aalborg University, Aalborg, Denmark
jonathansmets@gmail.com, jaeger@cs.aau.dk

Abstract. We propose a method for the simultaneous construction of multiple segmentations of images by combining a recently proposed “convolution of mixtures of Gaussians” model with a multi-layer hidden Markov random field structure. The resulting method constructs for a single image several, alternative segmentations that capture different structural elements of the image. We further introduce the notion of an image stack, by which we mean a collection of images with identical pixel dimensions. Here it turns out that the method is able to both identify groups of similar images in the stack, and to provide segmentations that represent the main structures in each group. We describe a variety of experimental results that illustrate the capabilities of the method.

Keywords: Segmentation · Multiple clustering · Probabilistic models

1 Introduction

Traditional clustering methods construct a single (possibly hierarchical) partitioning of the data. However, clustering when used as an explorative data analysis tool may not possess a single optimal solution that is characterized as the optimum of a unique underlying score function. Rather, there can be multiple distinct clusterings that each represent a meaningful view of the data. This observation has led to a recent research trend of developing methods for *multiple clustering* (or *multi-view clustering*). The general goal of these methods is to automatically construct several clusterings that represent alternative and complementary views of the data (see [12] for a recent overview, and the proceedings of the MultiClust workshop series for current developments).

The perhaps most typical application area for multiple clustering is document data (e.g. collections of news articles or web pages). For example, the standard benchmark WebKB dataset consists of university webpages that can be alternatively clustered according to page-type (e.g. personal homepage or course page), or the different universities the pages are taken from. Turning to image data, previously used benchmark sets are the CMU and the Yale Face Images data, which consists of portrait images of different persons in several poses, and accordingly can be clustered according to persons or poses [4, 8]. In this setting, each image is a data-point, and (multiple) clustering means grouping images. When, instead, one views as a data-point a single image pixel, then multiple clustering becomes *multiple image segmentation*.

Relatively little work has been done on finding multiple, alternative image segmentations. Reference [10] developed a quite specific *factorial Markov random field* model in which an image is modeled as an overlay of several layers, and each layer corresponds to a binary segmentation. Reference [14] apply a general multiple clustering approach to a variety of datasets, including images. Their multiple clustering approach falls into the category of iterative multiple clustering, where given an initial (primary) clustering, a single alternative clustering is constructed. Our approach, on the other hand, falls into the category of simultaneous multiple clustering methods, where an arbitrary number of different clusterings is constructed at the same time, and without any priority ordering among the clusterings. Finally, [9] generate alternative segmentations based on color and texture features, respectively. However, the objective here is not to provide different, alternative segmentations, but to combine the two segmentations into a single one.

It is worth emphasizing that multiple clustering in the sense here considered is different from the construction of *cluster ensembles* [17]. In the latter, numerous clusterings are built in order to overcome the convergence to only locally optimal solutions of clustering algorithms, and to construct out of a collection of clusterings a single consensus clustering. The multiple segmentations in the sense of [6, 16] are segmentation analogues of cluster ensembles, not of multiple clusterings in our sense.

In this paper we develop a method for constructing multiple segmentations of images and *image stacks*, which we define as a collection of images with equal pixel dimensions. The most import type of image stacks are the collection of frames in a video sequence. However, we can also consider other such collections of pixel-aligned images. As we will see in the experimental section, multiple clustering of such image stacks can give results that combine elements of clustering at the image and at the pixel level. For the design of our method we build on the *convolution of mixtures of Gaussians* model of [8] which we customize for the segmentation setting by combining it with a Markov Random Field structure to account for the spatial dimension of the data.

Our approach is intended as a general method that can be applied to image data of quite different types, and that thereby is a quite general tool for explorative image data analysis. For more specialized application tasks, our general method may serve as a basis, but will presumably require additional modifications and adaptations.

2 The Convolutional Clustering Model

Probabilistic clustering approaches are based on *latent variable models* where a data point $\mathbf{x}$ is assumed to be sampled from a joint distribution $P(\mathbf{X}, L \mid \boldsymbol{\theta})$ of an observed data variable $\mathbf{X}$ and a latent variable $L \in \{1, \dots, k\}$, governed by parameters $\boldsymbol{\theta}$ (throughout this paper we use bold symbols to denote tuples of variables, parameters, etc.; when talking about random variables, then upper-case letters stand for the variables, and lowercase letters for concrete values of the variables). Clustering then is performed by learning the parameters $\boldsymbol{\theta}$, and assigning $\mathbf{x}$ to the cluster with index i for which $P(\mathbf{X} = \mathbf{x}, L = i \mid \boldsymbol{\theta})$ is maximal.

Fig. 1. Multi-layer Hidden Markov Random Field.

This probabilistic paradigm is readily generalized to multiple clustering models. One only needs to design a model $P(\mathbf{X}, \mathbf{L} \mid \boldsymbol{\theta})$ containing multiple latent variables $\mathbf{L} = L_1, \dots, L_m$. Then the joint assignment $L_1 = i_1, \dots, L_m = i_m$ (abbreviated $\mathbf{L} = \mathbf{i}$) maximizing $P(\mathbf{X} = \mathbf{x}, L_1 = i_1, \dots, L_m = i_m \mid \boldsymbol{\theta})$ defines the cluster indices for $\mathbf{x}$ in m distinct clusterings. Models for multiple clustering that are based on multiple latent variables include the factorial Hidden Markov Model [5], the factorial Markov Random Fields of [10], convolution of mixtures of Gaussians [8], the latent tree models of [13], and the factorial logistic model of [7].

2.1 The Probabilistic Model

Our model is structurally identical to the factorial Markov Random Field model of [10]. Figure 1 shows the structure of such a *multi-layer hidden Markov random field*: with each pixel $i \in I$ (I the set of all pixels) are associated m latent variables $\mathbf{L}_{i,\bullet} = L_{i,1}, \dots, L_{i,m}$ and a vector of observed variables $\mathbf{X}_i$. For $k = 1, \dots, m$ the variables $\mathbf{L}_{\bullet,k} = L_{1,k}, \dots, L_{|I|,k}$ take values in the set $\{1, \dots, n_k\}$, so that the k th segmentation will consist of n_k segments.

For this paper we assume that in the case of single image analysis, $\mathbf{X}_i$ is simply the 3-dimensional vector (R_i, G_i, B_i) of *rgb*-values at pixel i . In the case of image stacks with N images, $\mathbf{X}_i$ will be a $3 \cdot N$ -dimensional vector containing the *rgb*-values of all images in the stack. We denote with $|\mathbf{X}|_i$ the dimension of $\mathbf{X}_i$. Though we do not explore this in the current paper, we note that $\mathbf{X}_i$ could also contain differently defined observed features of pixel i .

For every $k = 1, \dots, m$, the latent variables $\mathbf{L}_{\bullet,k}$ form a Markov random field with a square grid structure. The distribution of $\mathbf{X}_i$ depends conditionally on the latent variables $\mathbf{L}_{i,\bullet}$.

The marginal distribution $P(\mathbf{L} \mid \boldsymbol{\theta})$ is defined as a product of m Potts models defined by a common temperature parameter T :

$$P(\mathbf{L} = \mathbf{l} \mid \boldsymbol{\theta}) = P(\mathbf{L} = \mathbf{l} \mid T) = \frac{1}{Z} \prod_{k=1}^m e^{V(\mathbf{L}_{\bullet,k} = \mathbf{l}_{\bullet,k})/T}$$

where Z is the normalization constant, and

$$V(\mathbf{L}_{\bullet,k} = \mathbf{l}_{\bullet,k}) = \sum_{i,j:i \sim j} \mathbb{I}(l_{i,k} \neq l_{j,k})$$

with $\mathbb{I}(l_{i,k} \neq l_{j,k}) = 1$ if $l_{i,k} \neq l_{j,k}$, and $= 0$ otherwise.

For the conditional distribution $P(\mathbf{X} \mid \mathbf{L}, \boldsymbol{\theta})$ the model of Fig. 1 implies conditional independence for different pixels of the observed pixel features $\mathbf{X}_i$ given the latent pixel variables $\mathbf{L}_{i,\bullet}$. Moreover, we assume that the conditional model $P(\mathbf{X}_i \mid \mathbf{L}_{i,\bullet}, \boldsymbol{\theta})$ is identical for all i . It is defined as the convolution of m mixtures of Gaussians as follows. For $k = 1, \dots, m$ and $j = 1, \dots, n_k$ let $\mu_{k,j} \in \mathbb{R}^{|\mathbf{X}|}$. Writing $\boldsymbol{\mu}_k = \mu_{k,1}, \dots, \mu_{k,n_k}$, we obtain for every k a distribution for a variable $\mathbf{Z}_{i,k}$ defined as a mixture of Gaussians

$$P(\mathbf{Z}_{i,k} \mid L_{i,k}, \boldsymbol{\mu}_k) = \sum_{j=1}^{n_k} N(\mu_{k,j}, \mathbf{1}) \mathbb{I}(L_{i,k} = j),$$

where $\mathbf{1}$ stands for the unit covariance matrix. For two distributions $P(\mathbf{Y}), P(\mathbf{Z})$ of two k -dimensional real random variables $\mathbf{Y}, \mathbf{Z}$, we denote with $P(\mathbf{Y}) * P(\mathbf{Z})$ their convolution, i.e., the distribution of the sum $\mathbf{X} = \mathbf{Y} + \mathbf{Z}$. The final model for $\mathbf{X}_i$ now is defined as the m -fold convolution:

$$P(\mathbf{X}_i \mid \mathbf{L}_{i,\bullet}, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m) = P(\mathbf{Z}_{i,1} \mid L_{i,1}, \boldsymbol{\mu}_1) * \dots * P(\mathbf{Z}_{i,m} \mid L_{i,m}, \boldsymbol{\mu}_m).$$

Combining the model for $\mathbf{L}$ and $\mathbf{X} \mid \mathbf{L}$, We now obtain

$$\begin{aligned} \log(P(\mathbf{L} = \mathbf{l}, \mathbf{X} = \mathbf{x} \mid \boldsymbol{\mu}, T)) \approx \\ -1/T \sum_{k=1}^m \sum_{i,j:i \sim j} \mathbb{I}(l_{i,k} \neq l_{j,k}) - \sum_{i \in I} \left\| \mathbf{x}_i - \sum_{k=1}^m \mu_{k,l_{i,k}} \right\|^2 \quad (1) \end{aligned}$$

2.2 The Regularization Term

Maximizing the log-likelihood (1) alone is a sound approach to probabilistic multiple segmentation. However, [8] suggest to add to the likelihood the *regularization term*

$$- \lambda \sum_{\substack{k,k'=1,\dots,m \\ k \neq k'}} \sum_{\substack{j=1,\dots,n_k \\ j'=1,\dots,n_{k'}}} (\mu_{k,j} \cdot \mu_{k',j'})^2 \quad (2)$$

Here $\lambda \geq 0$ is a weight parameter that regulates the strength of the influence of the regularization term. This penalty term is minimized when the means $\boldsymbol{\mu}_k, \boldsymbol{\mu}_{k'}$ corresponding to different segmentations lie in orthogonal subspaces. The rationale given for this regularization term is twofold. First, the likelihood function (1) does not have a unique maximum. Indeed, taking the case $m = 2$, the two solutions $(\mu_{1,1}, \dots, \mu_{1,n_1}, \mu_{2,1}, \dots, \mu_{2,n_2}, T)$ and $(\mu_{1,1} + c, \dots, \mu_{1,n_1} + c, \mu_{2,1} - c, \dots, \mu_{2,n_2} - c, T)$ ($c \in \mathbb{R}^3$) define the same distribution, and therefore

have the same likelihood score. Second, the likelihood alone does not give an explicit reward for the distinctness, or complementarity, of the resulting multiple clusterings. Following other approaches to multiple clustering, it is hoped that encouraging the means corresponding to different clusterings to lie in orthogonal subspaces will lead to a greater diversity of those clusterings.

We argue that the form and justification for this particular regularization term is slightly flawed, and that it should be replaced by a modified version. First, we note that the non-uniqueness of the optimal solution for (1) is not a real problem as long as two different optimal solutions define the same multiple segmentation. This, however, is exactly the case for the two solutions distinguished by the offset vector c as described above. Second, regularization with (2) is not invariant under simple shifts of the coordinate system: adding a constant vector $\mathbf{z}$ to all data-points $\mathbf{x}_i$ should have no effect on the optimal segmentation, which should be characterized by also adding $\mathbf{z}$ to all model parameters $\mu_{k,j}$. Since (2) is not invariant under addition of a constant to all $\mu_{k,j}$, this is not the behavior one obtains with this regularization term. We therefore propose to modify (2) so as to reward means $\mu_k, \mu_{k'}$ to lie in orthogonal affine sub-spaces, rather than orthogonal linear sub-spaces. We therefore propose the following regularization term:

$$-\lambda \sum_{\substack{k,k'=1,\dots,m \\ k \neq k'}} \sum_{\substack{j,h=1,\dots,n_k:j < h \\ j',h'=1,\dots,n_{k'}:j' < h'}} \left(\frac{\mu_{k,j} - \mu_{k,h}}{\|\mu_{k,j} - \mu_{k,h}\|} \cdot \frac{\mu_{k',j'} - \mu_{k',h'}}{\|\mu_{k',j'} - \mu_{k',h'}\|} \right)^2. \quad (3)$$

Thus, we reward solutions in which normalized difference vectors between the means of different layers are orthogonal, rather than the means themselves. The term (3) now is invariant under adding, respectively subtracting, a constant vector c to all means of two different layers, and hence we again have the non-uniqueness of optimal solutions as for the pure likelihood (1). However, as argued above, we do not see this as a problem.

One small practical problem arises when we define our objective function as the sum of (1) and (3): the likelihood term (1) increases in magnitude linearly with the number of pixels. The regularization term, on the other hand, only increases as a function of the number of layers and the number of segments per layer. The choice of an appropriate tradeoff parameter λ between likelihood and regularization term, thus, would depend on the number of pixels. In order to get a more uniform scale for λ across different experiments, we therefore normalize the regularization term with the factor $|I|/K$, where K is the number of terms in the sum (3).

We remark that the probabilistic model (1) alone also has some built-in capability to encourage a diversity in the parameters μ_k for different layers, and hence, in the different segmentations. This is because having two layers with very similar means μ_k does not allow a much better fit to the data than a single layer with those means. Exploiting the full parameter space of the model to obtain a good fit to the data, thus, will tend to lead to some diversity in the parameters

μ_k . For this reason, in our experiments, we also pay particular attention to the case $\lambda = 0$, i.e., segmentation according to the pure probabilistic model (1).

The regularization terms (2) and (3) are intended to stimulate diversity in the computed segmentations, but they are not necessarily very meaningful, direct measurements for the diversity obtained. A common way to directly measure dissimilarity of two clusterings L_1, L_2 is *normalized mutual information*

$$NMI(L_1, L_2) = \frac{MI(L_1, L_2)}{\sqrt{H(L_1)H(L_2)}},$$

where MI is the mutual information and $H()$ the entropy of L_1, L_2 , as determined by the empirical joint distribution of L_1, L_2 defined by the cluster assignments of the pixels. Low values of NMI indicate statistical independence, and hence dissimilarity of clusterings. Furthermore, a justification given by [8] for the regularization term (2) is that it induces a bias towards statistically independent clusterings. This justification carries over to our modified version (3). Therefore, the NMI as an evaluation measure is quite consistent with our objective function.

2.3 Clustering Algorithm

We take the model parameter $\beta := 1/T$ and the regularization parameter λ as user-defined inputs that may be varied in an iterative data exploration process. Large values of β mean that high emphasis is put on segmentations with large connected segments and smooth boundaries. Larger values of λ mean that diversity of segmentations as measured by the regularization term (3) is more strictly enforced.

Thus, the only model-parameters we have to fit are the mean vectors μ_k . Our goal, then, is to maximize a score function $S(\mu_1, \dots, \mu_m, \mathbf{l})$ which is given as the sum of (1) and (3).

We use a typical 2-phase iterative process for this optimization: in a *MAP*-step we compute for a current setting of the μ_k the most probable assignment $\mathbf{L} = \mathbf{l}$ for the latent variables according to the likelihood function (1) (since (3) does not depend on $\mathbf{l}$, we can ignore it in this phase). In a *M(aximization)*-step we recompute for the current setting $\mathbf{L} = \mathbf{l}$ the μ_k optimizing $S(\mu_1, \dots, \mu_m, \mathbf{l})$. This well-known clustering approach (sometimes referred to as *hard EM*) has also been proposed for image segmentation in [3].

MAP-step. For the MAP-step we make use of the α -expansion algorithm of [1, 2, 11]. This algorithm provides solutions to segmentation problems characterized by an energy function E for segmentations s , which are of the form

$$E(s) = \sum_{i,j:i \sim j} V_{i,j}(s(i), s(j)) + \sum_i D_i(s(i)), \quad (4)$$

where $s(i)$ is the segment label of pixel i , $V_{i,j}$ is a penalty function for discontinuities in s , and D_i is any non-negative function measuring the discrepancy of

the label assignment $s(i)$ with the observed data for i . It is shown in [2] that if $V_{i,j}(s(i), s(j))$ is a metric on the label space, then the α -expansion algorithm is guaranteed to find a solution s that is within a constant factor of the globally minimal energy $E()$.

Up to a change of sign (and a corresponding change from a minimization to a maximization objective) our likelihood function (1) has the form (4) for the m -dimensional label space $\times_{k=1}^m \{1, \dots, n_k\}$ (i.e. $s(i) = (l_{i,1}, \dots, l_{i,m})$), with $V_{i,j}(s(i), s(j)) = \sum_{k=1}^m \mathbb{I}(l_{i,k} \neq l_{j,k})$ and $D_i(s(i)) = \|\mathbf{x}_i - \sum_{k=1}^m \mu_{k,l_{i,k}}\|^2$.

Furthermore, it is straightforward to see that our $V_{i,j}$ is a metric on the m -dimensional label space.

To use the α -expansion algorithm we flatten our m -dimensional label space to a one-dimensional label space with $\prod_{k=1}^m n_k$ different labels. Thus, our method has a complexity that is exponential in the number of layers. On the other hand, the α -expansion algorithm in practice is quite efficient as a function of the number of pixels. It is reputed to show a linear complexity in practice [2], which was confirmed by what we observed in our experiments.

M-step. The M-step is performed by gradient ascent, leading to a local maximum of the score function given the current segmentation $\mathbf{L} = \mathbf{l}$.

Implementation. The algorithm is implemented in Matlab, using the α -expansion implementation provided by the gco-v3.0 library¹.

3 Experiments

In all our experiments we construct multiple segmentations with the same number of segments in each layer. We therefore refer to a multiple segmentation with m layers and k segments in each layer as a (m, k) -segmentation.

3.1 Single Images

Our first experiment establishes the baseline result that the segmentation methods works as intended when the input closely fits the underlying modeling assumption. To this end we construct the image shown in Fig. 2(c) as the overlay of the two images (a) and (b), and used our method to construct (2,3)-segmentations from the single input image (c). First setting $\lambda = \beta = 0$, we performed 200 runs of the algorithm with different random initializations. The highest-scoring solution that was found consists of the segmentations (d) and (e). In these figures, the color of the j th segment in the k th layer is set to $\tilde{\mu}_{k,j}$, where $\tilde{\mu}_{k,j}$ is obtained from $\mu_{k,j}$ by applying min-max normalization to re-scale the components of all the mean vectors $\boldsymbol{\mu}_k$ ($k = 1, \dots, m$) into the interval [0..255] of proper rgb-values. Essentially the same optimal result was found in

¹ <http://vision.csd.uwo.ca/code/>.

Fig. 2. Baseline: overlay image. **Fig. 3.** Escher's butterflies: (a) original, (b) with added squares.

9 out of the 200 runs. In the remaining runs the algorithm converged to local maxima, an example of which is shown by (f) and (g). These results were clearly identified by the algorithm as sub-optimal by being associated with significantly lower score function values.

With increasing λ parameter the results in this experiment deteriorated. At $\lambda = 5000$ the “correct” solution was not found in 200 restarts. This is not very surprising, since for this image with $\lambda = \beta = 0$ the correct solution is clearly distinguished as the solution that can achieve a perfect score of 0 on the remaining Euclidean part of the likelihood term (1).

Next, we perform a series of experiments on the butterflies image by M.C. Escher, shown in Fig. 3(a), which has previously been used in [14]. The size of this image is 402×401 pixels.

We first compute (2,3)-segmentations with varying values of λ (and $\beta = 0$). Figure 4 shows the highest scoring results (in 20 restarts) obtained for $\lambda = 0, 1000, 10000$. In all cases, essentially the same two segmentations are computed: one that corresponds to the main colors of the three types of butterflies in the image, and one that captures the finer structure of the borders between the butterflies, as well as the shading inside the butterflies. The main effect of the regularization term here is not a difference in the segmentations, but only a difference in the means associated with the segments: for the high value $\lambda = 10000$, the means in the second segmentation all have a strong green component, whereas the means of the first component only have weak green components. This makes the means of the two components lie in near-orthogonal affine spaces. A similar color-separation does not appear at $\lambda = 0$.

Fig. 4. Escher (2,3)-segmentations, varying λ (Color figure online).

Fig. 5. Mutual information vs. complementarity.

As discussed in Sect. 2.2, the regularization term is intended to stimulate complementarity of segmentations, whereas NMI would be used to actually measure complementarity. In this experiment the increasing λ -values place a higher weight on the regularization term, and the value of the regularization term decreases from $8.28 \cdot 10^6$ for the solution at $\lambda = 1000$ to $1.82 \cdot 10^6$ at $\lambda = 10000$ (at $\lambda = 0$ no regularization term is computed). However, the NMI values for the three solutions of Fig. 4 are $8.4 \cdot 10^{-3}$, $5.4 \cdot 10^{-2}$, $7.1 \cdot 10^{-2}$ for $\lambda = 0, 1000, 10000$, respectively. Thus, the NMI values are even slightly increasing for larger λ -values.

We note at this point that NMI values have to be used with caution when assessing dissimilarity of image segmentations (rather than other types of data clusterings): NMI is a function only of cluster membership of pixels. However, for segmentations one is perhaps more interested in the borders defined between segments, than in the global grouping of pixels into segments. To illustrate this issue we consider the modified butterfly image in Fig. 3(b), in which we have superimposed an additional square grid structure on the original image. Figure 5(a) shows a hypothetical (2,4)-segmentation (not computed by our method) of this image. Both segmentations identify the grid structure – the first one dividing

the structure according to columns (and background), the second according to rows (and background). For the non-background pixels row and column membership are independent random variables. The mutual information of the two segmentations therefore reduces to $-P(b) \log P(b) - (1 - P(b)) \log(1 - P(b))$, where $P(b)$ is the probability of background pixels (i.e. the relative image area covered by background). In the limit where the size of the squares is increased, and $P(b) \rightarrow 0$, the mutual information of the two segmentations, thus, goes to zero (and so does the normalized mutual information). This shows that dissimilarity as measured by low mutual information need not correspond to the kind of complementarity we may be looking for in different segmentations. Figure 5(b) shows the (2,4)-segmentation actually obtained by our method. The result shown is for $\lambda = 0$, but results for higher λ -values are similar.

We conclude that neither need there be a good correspondence between low NMI values and complementarity of segmentations in the intuitive sense, nor does the regularization term necessarily induce a strong bias towards low NMI solutions. Fortunately, as Fig. 5(b) shows, the likelihood score alone is quite successful in producing segmentations that are complementary in an intuitively meaningful sense.

As a final experiment with the butterfly image, we do a (3,2)-segmentation with $\lambda = \beta = 0$. The result is shown in Fig. 6. The first segmentation again is based on the main underlying color distribution, isolating the blue butterflies

Fig. 6. Escher (3,2)-segmentation (Color figure online).

Fig. 7. Satellite image (Freiburg, Germany).

Fig. 8. Satellite: results with varying β .

from the rest. The last segmentation again represents mostly the border structure and shading. Finally, the segmentation in the middle is mostly identifying the green butterflies, but also represents some structure. Reference [14] present a (2,2)-segmentation for the butterfly image obtained from their iterative clustering method. Their two segmentations are quite similar in nature to the first two in Fig. 6.

We next use the satellite image shown in Fig. 7 to investigate the influence of the β -parameter, as well as the scalability properties of our method. Figure 8 shows the result of (2,2)-segmentations with $\beta = 500, 5000, 15000$. We first observe that in all cases one segmentation mostly singles out the valley/city region against the rest (top row), whereas the second segmentation distinguishes the wooded area (bottom row). Increasing β -values have their primary intended effect to produce more coherent segments with smoother boundaries. At the same time, with increasing β the complementarity of the two segmentations here becomes rather more pronounced, and the valley/city segment shrinks to a segment more specifically identifying the city areas only. All results presented here are the top scoring results out of 10 random restarts for each setting of β .

The input image used for the experiment shown in Fig. 8 had a resolution of $500 \times 346 = 173.000$ pixels. Figure 9 shows the runtime per restart for varying resolutions of the same input image. We clearly observe a linear scaling of the runtime as a function of the image size, which, in particular, confirms the in practice linear behavior of the α -expansion algorithm.

3.2 Image Stacks

As a first experiment with an image stack, we used the collection of 25 flag-images shown in Fig. 10 (each at a resolution of 150×75 pixels).

Again setting $\lambda = \beta = 0$, the highest scoring (2,3)-segmentation is shown at the right of Fig. 10. Here we now depict the different segments using arbitrarily

Fig. 9. Computation time for (2,2)-segmentations of Fig. 7.

Fig. 10. Stack of flag images (Color figure online).

Fig. 11. Stack of Horse and Train images.

chosen greyscale values. The means $\mu_{k,j}$ characterizing segments now are $3 \cdot 25$ dimensional vectors that can be interpreted as an average color sequence for pixels in a segment. Taking for visualization the average over all colors in the sequence typically leads to all segments represented by very similar brownish colors (although, curiously, in this particular case the average colors for the segmentation with the vertical stripes yield a somewhat washed-out looking French flag). The same “correct” solution here was found in 9 out of 50 random restarts.

A second image stack we constructed consists of 10 images each of trains and horses, as shown in Fig. 11. We performed (2,3)-segmentation with $\lambda = 0$ and $\beta = 50$. The highest scoring result within 400 runs is shown at the right of Fig. 11. The method identifies the main structures in the two groups of images also in this somewhat more diverse collection of images. The results in the different runs were relatively stable, with other high-scoring solutions similar to the top-scoring one. Results with lower scores often separated the two groups of images less clearly, or contained segmentations in which one segment was reduced to very few pixels. The average runtime per restart in this experiment was about 1 min.

Fig. 12. Weather satellite image stack.

In a last image stack experiment we use a stack of 17 weather satellite images showing the cloud distribution over Europe on different days in the summer months June-August in years 2011–2014². Figure 12 shows a representative 4 of the 17 input images, and the highest scoring result from 10 restarts of a (2,2)-segmentation. Interestingly, the top 5 solutions in the 10 restarts were visually indistinguishable from the one shown in Fig. 12, and achieved almost the same optimal score (note that even identical segmentations can have somewhat different scores, because the score is a function of the underlying model parameters $\mu_{k,j}$, not the segmentation alone). This robustness in the results under random restarts indicates that the found (2,2)-segmentation really shows relevant patterns in the input data, which one may cautiously try to interpret as patterns of cloud distributions.

² Image source: <http://www.sat24.com>.

In all our experiments results were quite robust under variations of the λ and β parameters. Good results are typically already obtained at the baseline setting $\lambda = \beta = 0$. Note that $\beta = 0$ means that the Markov random field structure of the model is ignored, and that the MAP step could be implemented in a much simplified manner. In applications where smooth and contiguous segments are required, settings of $\beta > 0$ will be needed. The impact of the λ parameter on the segmentations was rather small. It appears that larger values of λ affected the placement of the mean parameters representing the different segments, but not so much the resulting segmentations themselves.

4 Conclusions

We have introduced a method for constructing multiple segmentations of image stacks by combining the convolution of mixtures of Gaussians model [8] with a multi-layer Markov Random field. While novel in this form, the resulting model is a quite straightforward combination of existing components. The main original contribution of this paper is the first dedicated investigation of multiple clustering for image segmentation, and the introduction of (multiple) segmentation of image stacks. We note that the latter is different from cosegmentation [15] and standard video segmentation, where also “stacks” of images are segmented simultaneously, but where a separate segmentation is computed for each image (or frame).

We have conducted a range of experiments that demonstrate that the method is able to produce meaningful results in a broad variety of datasets. Applied to single images, it is able to identify the structures of multiple constituent components. Applied to image stacks, it can perform a simultaneous clustering at the image and at the pixel level. All these results were obtained using only the basic rgb pixel features. No task-specific preprocessing or feature engineering was needed to obtain our results. One can thus conclude, that the proposed method provides a useful baseline approach for explorative image analysis.

For more specific application purposes or data analysis objectives, it will be necessary to construct more specific pixel features. One possible such application domain is multiple segmentation of video sequences. The frames of a video can obviously be seen as an image stack. Using only the rgb pixel features our method is not very well adapted to video analysis, since it does not take into account the temporal order of the frames. New pixel features that capture some of the temporal dynamics of the pixel values can be constructed, for example, simply by considering the variance of the pixel’s rgb values, or by constructing features that describe the trajectory of the pixel’s rgb values in rgb-space. Performing multiple segmentation of video sequences based on such features is a topic for future work.

In this paper we have also tried to evaluate the usefulness of regularization terms along the lines proposed in [8] for stimulating diversity in the multiple segmentations. Our results lead to some doubts both with regard to the effectiveness

of the regularization term to produce segmentations with low mutual information, and with regard of the usefulness of mutual information as a measure for diversity in image segmentations. On the other hand, our results indicate that the likelihood term (1) alone is quite capable of identifying the most relevant, distinct segmentations.

References

1. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(9), 1124–1137 (2004)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1222–1239 (2001)
3. Chen, S., Cao, L., Wang, Y., Liu, J., Tang, X.: Image segmentation by MAP-ML estimations. *IEEE Trans. Image Process.* **19**(9), 2254–2264 (2010)
4. Cui, Y., Fern, X., Dy, J.: Non-redundant multi-view clustering via orthogonalization. In: *Proceedings of Seventh IEEE International Conference on Data-Mining (ICDM 2007)*, pp. 133–142 (2007)
5. Ghahramani, Z., Jordan, M.: Factorial hidden markov models. *Mach. Learn.* **29**(2–3), 245–273 (1997)
6. Hoiem, D., Efros, A., Hebert, M.: Geometric context from a single image. In: *Tenth IEEE International Conference on Computer Vision (ICCV 2005)*, pp. 654–661 (2005)
7. Jaeger, M., Lyager, S.P., Vandborg, M.W., Wohlgenuth, T.: Factorial clustering with an application to plant distribution data. In: *Proceedings of the 2nd Multi-Clust Workshop*, pp. 31–42 (2011). Online proceedings <http://dme.rwth-aachen.de/en/MultiClust2011>
8. Jain, P., Meka, R., Dhillon, I.S.: Simultaneous unsupervised learning of disparate clusterings. *Stat. Anal. Data Min.* **1**(3), 195–210 (2008)
9. Kato, Z., Pong, T.-C., Qiang, S.G.: Unsupervised segmentation of color textured images using a multilayer MRF model. In: *Proceedings of the IEEE International Conference on Image Processing (ICIP 2003)*, vol. 1, pp. 961–964. IEEE (2003)
10. Zabih, R., Kim, J.: Factorial Markov random fields. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) *ECCV 2002, Part III. LNCS*, vol. 2352, pp. 321–334. Springer, Heidelberg (2002)
11. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(2), 147–159 (2004)
12. Müller, E., Günnemann, S., Färber, I., Seidl, T.: Discovering multiple clustering solutions: grouping objects in different views of the data. In: *Proceedings of 28th International Conference on Data Engineering (ICDE 2012)*, pp. 1207–1210 (2012)
13. Poon, L.K.M., Zhang, N.L., Chen, T., Wang, Y.: Variable selection in model-based clustering: To do or to facilitate. In: *Proceedings of the 27th International Conference on Machine Learning (ICML 2010)*, pp. 887–894 (2010)
14. Qi, Z., Davidson, I.: A principled and flexible framework for finding alternative clusterings. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009)*, pp. 717–725 (2009)

15. Rother, C., Minka, T., Blake, A., Kolmogorov, V.: Cosegmentation of image pairs by histogram matching-incorporating a global constraint into MRFs. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 993–1000. IEEE (2006)
16. Russell, B., Freeman, W., Efros, A., Sivic, J., Zisserman, A.: Using multiple segmentations to discover objects and their extent in image collections. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1605–1614 (2006)
17. Strehl, A., Ghosh, J.: Cluster ensembles – a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.* **3**, 583–617 (2003)