A Spatially-Constrained Mixture Model for Image Segmentation

K. Blekas, A. Likas, N. P. Galatsanos*and I. E. Lagaris

Department of Computer Science, University of Ioannina 45110 Ioannina, Greece

E-mail: {kblekas,arly,galatsanos,lagaris}@cs.uoi.gr

Abstract

Gaussian Mixture Models (GMMs) constitute a well-known type of probabilistic neural networks. One of their many successful applications is in image segmentation, where spatially constrained mixture models have been trained using the Expectation-Maximization (EM) framework. In this paper, we elaborate on this method and propose a new methodology for the M-step of the EM algorithm that is based on a novel constrained optimization formulation. Numerical experiments using simulated images illustrate the superior performance of our method in terms of the attained maximum value of the objective function and segmentation accuracy compared to previous implementations of this approach.

1 Introduction

Image segmentation is the process that groups image pixels together based on attributes such as their intensity and spatial location. A variety of different methods have been proposed for image segmentation such as edge-based segmentation, region-based segmentation, pixel labeling and hybrid techniques [1, 2, 3]. In this work, we elaborate on a pixel labeling (clustering) technique based on Gaussian Mixture Models (GMM) which constitute a well-known probabilistic neural network model [4, 5]. The Expectation-Maximization framework constitutes an efficient method for GMM training based on likelihood maximization.

The application of clustering methods to image segmentation has the particular characteristic that spatial information should be taken into account. That is, apart from the intensity values, the pixel location must also be used to determine the cluster to which each pixel is assigned. Intuitively speaking, in most cases it is desirable to assign the same cluster label to spatially adjacent pixels. The Bayesian framework provides a natural approach to implement these ideas. Following this formulation, a likelihood term which is based exclusively on the data captures the pixel intensity information, while a prior biasing term that uses a Markov Random Field (MRF) captures the spatial location information. Thus, it is no surprise that most recent image segmentation algorithms follow this paradigm; see for example [6, 7].

Nevertheless, an inherent difficulty with this formulation is that, due to the introduction of the prior, the M-step of the EM algorithm cannot be implemented using closed form expressions. For this reason, in [6], a Gradient Projection (GP) algorithm was proposed to implement the M-step.

---

*To whom correspondence should be addressed.

In this paper we propose a novel method to implement the M-step based on a closed form update equation followed by an efficient projection method. We demonstrate with numerical experiments using the synthetic image data in [7] that the proposed M-step provides a better maximum of the objective function than the GP approach proposed in [6]. In addition, it also yields better segmentation results.

The rest of this paper is organized as follows: In section 2 we describe the probabilistic model for image segmentation. In section 3 we present our improvements to this model. In section 4 we provide comparative experimental results and finally in section 5 our conclusions and future work.

2   The spatially variant finite mixture model

Let x^i denote the observation at the ith pixel of an image ($i = 1, \ldots, N$) modeled as i.i.d. The spatially variant finite mixture model (SVFMM) [6] provides a modification of the classical mixture model approach for pixel labeling. The SVFMM assumes a mixture model with K components each one having its own vector of density parameters θ^j.

According to the SVFMM approach, the probabilities $\pi_j^i = P(j|x^i)$ of the ith pixel belonging to the jth cluster (class label) are considered as additional model parameters that should satisfy the following constraints: $0 \leq \pi_j^i \leq 1$ and $\sum_{j=1}^{K} \pi_j^i = 1$. Let π^i denotes the probability vector for pixel i, $\Pi = \{\pi^1, \ldots, \pi^N\}$ the set of probability vectors and $\Theta = \{\theta^1, \ldots, \theta^K\}$ the set of component parameters. Then the SVFMM model assumes that the density function $f(x^i|\Pi, \Theta)$ at an observation x^i is given by

$$f(x^i|\Pi, \Theta) = \sum_{j=1}^{K} \pi_j^i \phi(x^i|\theta^j) \ , \tag{1}$$

where $\phi(x^i|\theta^j)$ is a Gaussian distribution with parameters $\theta^j = \{\mu_j, \sigma_j\}$.

Based on the above formulation the parameters of the model can be estimated through likelihood maximization (ML) using the EM algorithm. Since the pixel observations are considered to be independent samples, a significant drawback of the ML approach is that the spatial pixel information is not taken into account [8, 7]. To overcome this difficulty the SVFMM method considers a *maximum a posteriori* (MAP) approach by introducing a prior distribution for the parameter set Π that takes into account spatial information based on the following Gibbs function [8, 6, 7]

$$p(\Pi) = \frac{1}{Z} \exp(-U(\Pi)) \ , \text{ where } U(\Pi) = \beta \sum_{i=1}^{N} V_{\mathcal{N}_i}(\Pi) \ . \tag{2}$$

The Z is a normalizing constant, while β is oftenly called regularization parameter. The function $V_{\mathcal{N}_i}(\Pi)$ denotes the clique potential function of the pixel label vectors $\{\pi^m\}$ within the neighborhood $\mathcal{N}_i$ of the ith-pixel and can be computed as follows

$$V_{\mathcal{N}_i}(\Pi) = \sum_{m \in \mathcal{N}_i} g(u_{i,m}) \ , \tag{3}$$

where the $u_{i,m}$ specifies the distance between the two label vectors π^i and π^m, i.e, $u_{i,m} = |\pi^i - \pi^m|^2 = \sum_{j=1}^{K}(\pi_j^i - \pi_j^m)^2$. The neighborhood $\mathcal{N}_i$ is the set containing pixels that are horizontally or vertically adjacent to pixel i. Finally, the function $g(u)$ must be nonnegative and monotonically increasing [8]. We have selected $g(u) = (1 + u^{-1})^{-1}$ adopted from [8], while in [6] the identity function $h(u) = u$ was used. The function $g(u)$ penalizes less large values of u and thus is more robust to outliers.

Given the above prior density, a posteriori log-density function can be as follows

$$p(\Pi, \Theta | X) = \sum_{i=1}^{N} \log f(x^i | \Pi, \Theta) + \log p(\Pi) . \tag{4}$$

The use of the EM algorithm for MAP estimation of the parameters $\{\pi_j^i\}$ and $\{\theta^j\}$ [6] requires that the conditional expectation values z_j^i of the hidden variables are computed at the E-step

$$z_j^{i^{(t)}} = \frac{\pi_j^{i^{(t)}} \phi(x^i | \theta_j^{(t)})}{\sum_{l=1}^{K} \pi_l^{i^{(t)}} \phi(\mathbf{x}^i | \theta_l^{(t)})} , \tag{5}$$

while in the M-step the maximization of the following log-likelihood corresponding to the complete data set is performed

$$Q_{MAP}(\Pi, \Theta | \Pi^{(t)} \Theta^{(t)}) = \sum_{i=1}^{N} \sum_{j=1}^{K} z_j^{i^{(t)}} \{ \log(\pi_j^i) + \log(\phi(x^i | \theta^j)) \} - \beta \sum_{i=1}^{N} \sum_{m \in \mathcal{N}_i} g(u_{i,m}) , \tag{6}$$

where t indicates the iteration step. The function Q_{MAP} can be maximized independently for each parameter. This gives the following update equations for parameters of the component densities

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_j^{i^{(t)}} x^i}{\sum_{i=1}^{N} z_j^{i^{(t)}}} \text{ and } [\sigma_j^2]^{(t+1)} = \frac{\sum_{i=1}^{N} z_j^{i^{(t)}} [x^i - \mu_j^{(t+1)}]^2}{\sum_{i=1}^{N} z_j^{i^{(t)}}} . \tag{7}$$

However, the maximization of the function Q_{MAP} with respect to the label parameters $\{\pi_j^i\}$ does not provide closed form update equations. In addition, the maximization procedure must also take into account the constraints $0 \leq \pi_j^i \leq 1$ and $\sum_{j=1}^{K} \pi_j^i = 1$. Due to this difficulty a Generalized EM scheme for estimating the label parameters $\{\pi_j^i\}$ was adopted in [6] following the iterative *Gradient Projection* method. According to this method the gradient of the MAP function is first projected onto the hyperplane of the constraints. Then a line search is performed along the direction of the projected gradient to find the label parameters $\{\pi_j^i\}$ that maximizes the Q_{MAP} function.

3   The proposed technique

In this section we present the new M-step which we demonstrate experimentally in section 4 that improves the performance of the segmentation algorithm. In order to maximize Q_{MAP}

(Eq. (6)) with respect π_j^i we set its derivative equal to zero and obtain the following quadratic expression

$$4\beta\left[\sum_{m\in\mathcal{N}_i}\dot{g}(u_{i,m})\right](\pi_j^{i^{(t+1)}})^2 - 4\beta\left[\sum_{m\in\mathcal{N}_i}\dot{g}(u_{i,m})\pi_j^m\right](\pi_j^{i^{(t+1)}}) - z_j^{i^{(t)}} = 0 , \tag{8}$$

where $\dot{g}(u)$ indicates the derivative of g. It must be noted that in the above equation the neighborhood $\mathcal{N}_i$ can include pixels with updated label parameter vectors ($t+1$ step), as well as pixels whose label vectors π^m have not yet been updated (t step).

The two roots of the above equation are

$$\pi_j^{i^{(t+1)}} = \frac{\left[\sum_{m\in\mathcal{N}_i}\dot{g}(u_{i,m})\pi_j^m\right] \pm \sqrt{\left[\sum_{m\in\mathcal{N}_i}\dot{g}(u_{i,m})\pi_j^m\right]^2 + \frac{1}{\beta}z_j^{i^{(t)}}\left[\sum_{m\in\mathcal{N}_i}\dot{g}(u_{i,m})\right]}}{2\left[\sum_{m\in\mathcal{N}_i}\dot{g}(u_{i,m})\right]} , \tag{9}$$

and select only the root with the positive sign $+$ since it yields $\pi_j^i \geq 0$. The above equation provides a straightforward update for the values of label parameters π_j^i of each pixel i at the M-step of every EM iteration. However, we also have to ensure that these values satisfy the constraints $0 \leq \pi_j^i \leq 1$ and $\sum_{j=1}^K \pi_j^i = 1$. In the following we present an efficient novel projection algorithm to achieve this goal.

For convenience, let us now denote with a_j $(j = 1,\dots,K)$ the label parameter values $(\pi_j^{i^{(t+1)}} \geq 0)$ computed from Eq. (9). The problem we address here is the following: "Given a vector $a \in \mathcal{R}^K$ with $a_j \geq 0$ and the hyperplane $\sum_{j=1}^K y_j = 1$, find the point y on the hyperplane with non-negative components that is closest to a". This can be formulated as a *linear constrained convex quadratic programming* (QP) problem:

$$\min_y \sum_{j=1}^K (y_j - a_j)^2$$
$$\text{subject to } \sum_{j=1}^K y_j = 1 \text{ and } y_j \geq 0 , \forall j = 1, 2, \dots, K . \tag{10}$$

In order to solve the above QP problem, several approaches may be employed [9], such as *active-set* methods that use Lagrange multiplies, as well as *penalty-barrier* methods that formulate an objective function with penalty terms for equality and barrier terms for inequality constraints. We use here an active-set type of method where we exploit the fact that the Hessian is the identity matrix which in turn leads to the derivation of closed form analytical expressions for the Lagrange multipliers. This is of great value for both the efficiency and the robustness of the method, since it avoids the burden of numerical instabilities that occur frequently in the solution of large linear systems when the associated matrices are nearly singular.

One may proceed using the following Lagrange function:

$$L(y, \lambda_0, \lambda) = \frac{1}{2}\sum_{j=1}^K (y_j - a_j)^2 - \lambda_0(\sum_{j=1}^K y_j - 1) - \sum_{j=1}^K \lambda_j y_j , \tag{11}$$

where λ_0 is the multiplier for the equality and λ_j, $j = 1, \cdots, K$ the multipliers for the inequality constraints. First order necessary conditions imply:

$$y_j = a_j + \lambda_0 + \lambda_j \ . \tag{12}$$

Combining the above with the equality constraint yields:

$$\lambda_0 = \frac{1}{K} - <a> - <\lambda> \ , \tag{13}$$

where $<v> \equiv \frac{1}{K} \sum_{j=1}^{K} v_j$. Hence substituting λ_0 in Eq. (12) we have that:

$$y_j = \frac{1}{K} + a_j - <a> + \lambda_j - <\lambda>, \quad j = 1, \cdots, K \ . \tag{14}$$

Note that the vector b with components $b_j = \frac{1}{K} + a_j - <a>$ is the projection of a on the hyperplane $\sum_{j=1}^{K} y_j = 1$. The λ's must be chosen so as to satisfy the inequality constraints. Khun-Tucker conditions [9] state that at the minimizer y^*:

$$\lambda_j \geq 0 \ , \quad \lambda_j > 0 \ \text{if} \ y_j^* = 0 \ (\text{Active constraint}) \ , \quad \lambda_j y_j^* = 0 \ . \tag{15}$$

We present a very efficient iterative strategy for calculating the λ's for the problem above.

Let y denote the vector at the current iteration. Initially we set $y_j = b_j$, $\forall j = 1, \cdots, K$. In the general case there exist m negative components y_j. The corresponding set of indices $S = \{j, \text{ with } y_j < 0\}$ constitutes the active set of constraints for the current vector y.

- For all $j \notin S$ we set $\lambda_j = 0$.

- For all $j \in S$ we set $y_j = y_j^\star = 0$ and we compute the corresponding λ_j by solving an $m \times m$ linear system that force the inequalities to be satisfied as equalities, namely $y_j + \lambda_j - <\lambda> = 0$, leading to

$$\lambda_j = \frac{1}{m - K} \sum_{k \in S} y_k - y_j \ . \tag{16}$$

- We compute the updated y_j values for $j \notin S$ using the new vector λ via Eq. (14).

The above procedure is repeated until a feasible point is obtained, i.e. $y_j \geq 0$, $\forall j$. This is the desired minimizer ($y^\star = y$).

Note that Eq. (16) produces positive values for λ_j, hence no constraint is to be dropped ever from the active set, ie. if once some y_j becomes zero then it retains this value for ever. This is a very important point as far as efficiency is concerned and in addition guarantees the finite termination property of the algorithm. When all constraints are satisfied we have reached the sought solution.

Finally, it must be noted that apart from problems of low dimensionality as in the case of image segmentation (where K is equal to the number of pixel labels), we have also applied the above projection technique to problems of large dimensions (eg. $K = 65536$) arising in other image processing problems and we have verified its efficiency and robustness.
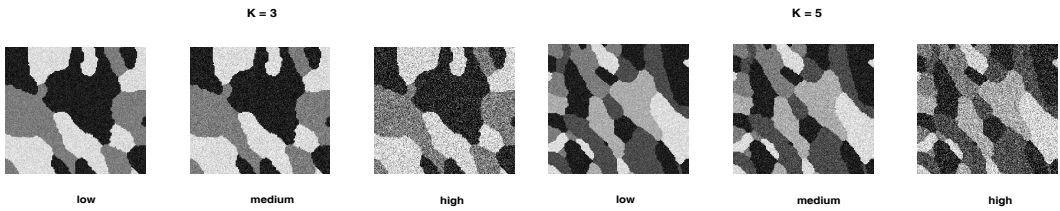
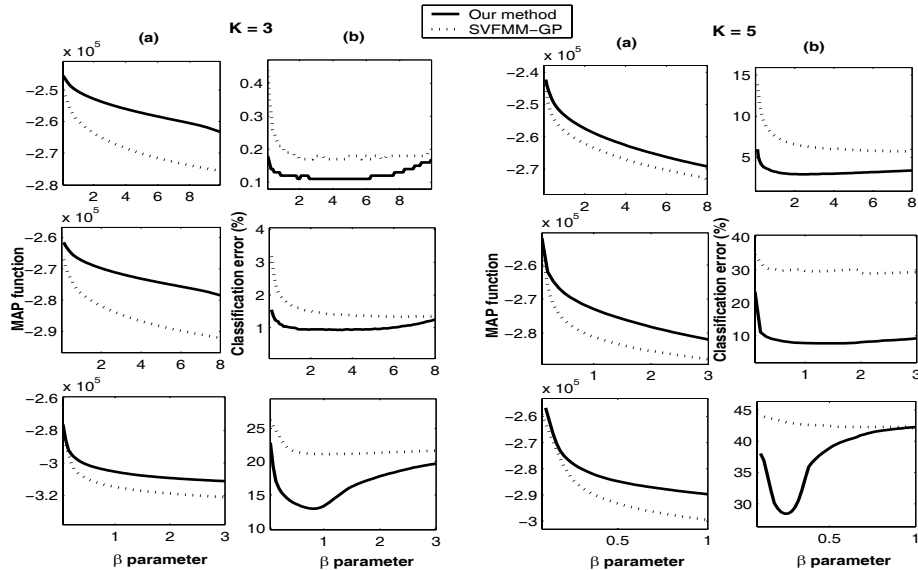Figure 1: Six noisy test images with 3 and 5 classes using three levels of noise.

Figure 2: Plot of the MAP function (a) and the classification error (b) for various β values in the case of the three noisy images with $K = 3$ and $K = 5$ classes.

4   Experimental results

A series of image segmentation experiments have been conducted to evaluate and compare the effectiveness of the proposed technique. Since the main contribution of our work is on improving the M-step of the SVFMM model that estimates the label parameters π_j^i, we compared our approach with the Generalized EM scheme proposed in the original SVFMM model description that employs the Gradient Projection technique (termed as SVFMM-GP) as described in [6].

In this paper we present results using two simulated test images being sampled from MRF model using a Gibbs sampler [7], with $K = 3$ and $K = 5$ classes, where we have added three levels of Gaussian noise with standard deviation of 18, 25 and 52, respectively (Fig. 1). Figure 2 illustrates the comparative results from the application of the two methods to each noisy image. Two evaluation criteria have been used for the comparison study: a) the maximum attained value of the MAP objective function (Eq. (4)) and b) the classification (segmentation) error defined as the percentage of mis-classified pixels. Therefore, for each image segmentation problem we provide two diagrams that illustrate the performance of the models according to the above two criteria for several values of the β parameter. These results, demonstrate that our approach provides a better maximum of the MAP function. Moreover, it provides significantly better
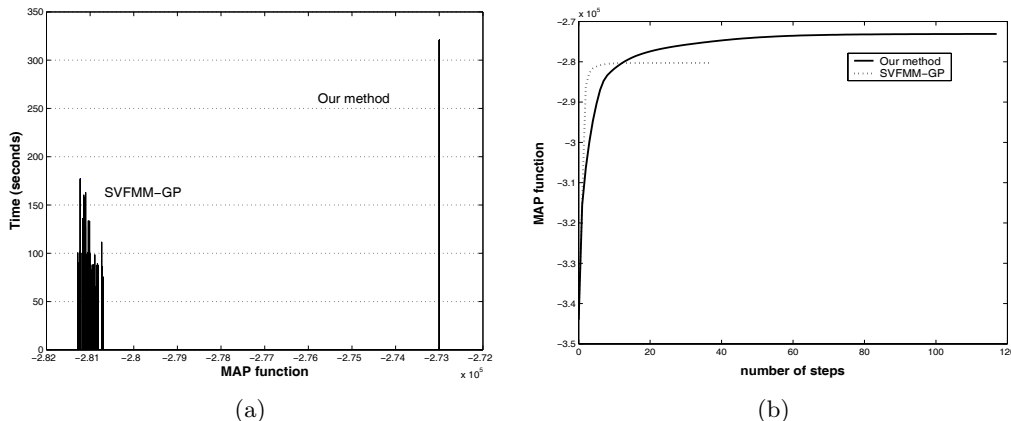
6

Figure 3: (a) Execution time versus MAP function values for 50 runs of the proposed and the SVFMM-GP methods. (b) A typical run example of MAP function versus number of steps is also presented.

segmentation accuracy, since the mis-classification ratio is considerably lower in our approach, especially for high levels of noise.

In order to gain insight on the reason why our method yields consistently better results, the following additional experiments were conducted For each method and test image (Fig. 1), we performed 50 runs starting from different initial points π_j^i while keeping the same initial values for the mixture component parameters (μ_j, σ_j). The regularization parameter β was kept equal to $\beta = 1$ throughout. In Fig. 3a we plot the execution time for each of the 50 runs, versus the obtained MAP function value using both methods on the fifth test image. Note that the results for all images were similar.

In Fig. 3 we observe that our method returns in all cases the same result, whereas the SVFMM-GP technique depends on the initialization, hence returning not one, but a number of different maxima. This behaviour can be explained noting that our method locates the global maximum of the Q_{MAP} function (Eq. (6)) in every M-step, since it only projects the unique unconstrained maximum (Eq. (9)) onto the constraint boundary. On the contrary, SVFMM-GP is trapped to the first local maximum encountered. There is a small variation in the execution times (≈ 1 sec) of our method that cannot be shown in Fig. 3a since the value of function is the same in all runs. This is expected since the projections onto the constraint boundary may require different times, depending on the initial conditions. However, in our exeriments the dimension of the constraint polyedron is small ($K = 3$ or 5), thus, these differences are negligible.

In Fig. 3b we plot the MAP function values versus the number of EM-steps. From that plot we see that our method requires more EM iterations to converge and it is slower as it can be deduced from Fig. 3a. However, our method reaches higher maximum values which implies better segmentation performance.

7

5 Conclusions

We have presented a new method to maximize the label parameter values at the M-step of the EM algorithm for training GMMs with MRF priors for image segmentation. Experimental results on simulated images demonstrate that the proposed modification improves, in some cases significantly, the segmentation performance of this method. It must be noted that we have also tested the proposed algorithm with real images, where in all tested cases the proposed M-step provides a better maximum of the objective function. However, due to space constraints, we do not present these results here. Future work will focus on applying the method to real world segmentation problems arising in medical imaging and bioinformatics and also to consider segmentation of color and texture images. We also plan to design more sophisticated prior functions that will take into account not only pixel adjacency, but also image information, such as for example the existence of edges.

References

[1] T. Pappas, "An Adaptive Clustering Algorithm for Image Segmentation," *IEEE Trans. on Signal Processing*, vol. 40, no. 4, pp. 901–914, 1992.

[2] N. Pal and S. Pal, "A Review of Image Segmentation Techniques," *Pattern Recognition*, vol. 26, pp. 1277–1294, 1993.

[3] K. Chen, D. Wang, and X. Liu, "Weight Adaptation and Oscillatory Correlation for Image Segmentation," *IEEE Trans. on Neural Networks*, vol. 11, no. 5, pp. 1106–1123, 2000.

[4] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford Univ. Press Inc., New York, 1995.

[5] D. Ormoneit and V. Tresp, "Averaging, Maximum Penalized Likelihood and Bayesian Estimation for Improving Gaussian Mixture Probability Density Estimates," *IEEE Trans. on Neural Networks*, vol. 9, no. 4, pp. 639–650, 1998.

[6] S. Sanjay-Gopal and T. J. Hebert, "Bayesian Pixel Classification Using Spatially Variant Finite Mixtures and the Generalized EM Algorithm," *IEEE Trans. on Image Processing*, vol. 7, no. 7, pp. 1014–1028, 1998.

[7] Y. Zhang, M. Brady, and S. Smith, "Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm," *IEEE Trans. on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.

[8] P. J. Green, "Bayesian Reconstructions from Emission Tomography Data Using a Modified EM Algorithm," *IEEE Trans. on Medical Imaging*, vol. 9, no. 1, pp. 84–93, 1990.

[9] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer-Verlag, New York, 1999.