BILATERAL FILTER BASED MIXTURE MODEL FOR IMAGE SEGMENTATION

Dibyendu Mukherjee, Q.M. Jonathan Wu, Thanh Minh Nguyen

Department of Electrical and Computer Engineering, University of Windsor
401 Sunset Avenue, Windsor, ON, N9B3P4, Canada
{mukherjd, jwu, nguyen1j}@uwindsor.ca

ABSTRACT

This paper introduces a bilateral filtering based mixture model for image segmentation. The mixture model uses Markov Random Field (MRF) to incorporate spatial relationship among neighboring pixels into the Gaussian Mixture Model (GMM) in order to perform a segmentation that is robust against noise and other environmental factors. The bilateral filtering is used to smooth the posterior probability map as part of the MRF used. The advantage of the proposed model is its simplified structure so that the Expectation Maximization algorithm can be directly applied to the log-likelihood function to compute the optimum parameters of the mixture model. The method has been extensively tested on synthetic and natural images and compared with some of the state-of-the-arts algorithms currently available. The experimental results show that the proposed method is comparable to the other methods in terms of accuracy and quality and simpler in terms of implementation.

Index Terms— Image segmentation, Gaussian mixture model, Markov random field, Bilatering Filtering, spatial information, EM algorithm.

1. INTRODUCTION

In computer vision related problems, image analysis is the most important part. In order to analyze an image, it is often more meaningful to partition the image into non-overlapping regions that correspond to some real world objects, to draw inference on the contents of the image. The process involves clustering pixels based on their intensity and spatial locations. Often, pixels corresponding to a single object share similar intensity values and neighboring relationship. These properties are exploited in segmentation. However, with increased amount of noise, it gets harder to achieve accurate image segmentation. In literature, there are many popular segmentation algorithms like mean shift based methods [1], graph based methods [2, 3], clustering approaches [4] etc. In recent years, Bayesian framework based approaches [5] have gained popularity due to their simplicity and ease of implementation. Among these methods, standard GMM [6, 7] is well known. One of the main drawbacks of this method is that the prior distribution π_j does not depend on the pixel index j and thus, on the spatial relationships between the labels of neighboring pixels. Thus, the segmentation is extremely noise-prone and illumination dependent. To overcome this disadvantage, mixture models with MRF have been employed for pixel labeling [8, 9]. The distinct difference is that the prior distribution π_{ij} varies for every pixel x_i corresponding to each label Φ_j and depends on the neighboring pixels and the corresponding parameters. The disadvantages of the MRF methods lie in lacking robustness against high amount of noise and increase in computational cost. Several researchers have extended the models [10, 11] where an MRF models the joint distribution of the priors of each pixel, instead of the joint distribution of the pixel labels.

In this work, a new MRF based mixture model is proposed. The model is made based on the following considerations - firstly, the model is very simple compared to the other MRF based models. The structure of the model has been reduced to simple filtering in probability domain. Secondly, the spatial information has been successfully incorporated in the model with the use of bilateral filtering and the EM-algorithm can be directly applied to compute the parameters of the method.

The paper has been organized in following sections. In section 2, the background of the current work is briefly discussed. The proposed method is described in section 3. Section 4 provides brief understanding of the Bilateral filtering used for the work. Section 5 includes the experimental results and finally, the work is concluded in section 6.

2. MIXTURE MODEL BASED ON MARKOV RANDOM FIELD

Let, x_i, $i = (1, 2, ..., N)$, where each x_i is of dimension D, denote an observation at the i^{th} pixel of an image. The neighborhood of the i^{th} pixel is presented by δ_i. The target is to associate each x_i with a label in $(1, 2, ..., K)$. For this classification, standard GMM assumes that each observation x_i is independent of the label Ω_j. The density function $f(x_i \mid \Pi, \Theta)$ at an observation x_i is given by:

$$f(x_i \mid \Pi, \Theta) = \sum_{j=1}^{K} \pi_{ij} \Phi(x_i \mid \Theta_j) \tag{1}$$

where, $\Pi = \{\pi_{ij}\}$, $j = (1, 2, ..., K)$ is the set of prior distributions of probabilities where π_{ij} denotes the probability that pixel x_i is in label Ω_j and satisfies the constraints:

$$0 \leq \pi_{ij} \leq 1 \text{ and } \sum_{j=1}^{K} \pi_{ij} = 1 \tag{2}$$

Also, $\Phi(x_i \mid \Theta_j)$ is a component of the gaussian mixture. Each component can be written in the form:

$$\Phi(x_i \mid \Theta_j) = \frac{|\Sigma_j|^{-1/2}}{(2\pi)^{D/2}} \exp\left\{-\frac{\Delta^2}{2}\right\} \tag{3}$$

where $\Delta^2 = (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)$ is the squared Mahalanobis distance and $\Theta_j = \{\mu_j, \Sigma_j\}$, $j = (1, 2, ..., K)$. The D-dimensional vector μ_j is the mean, the $D \times D$ matrix Σ_j is the covariance, and

This research has been supported in part by the Canada Research Chair Program, AUTO21 NCE, and the NSERC Discovery grant.

$|\Sigma_j|$ denotes the determinant of Σ_j . From Eq.(1), the joint conditional density of the data set $X = (x_1, x_2, ..., x_N)$ can be written as:

$$p(X \mid \Pi, \Theta) = \prod_{i=1}^{N} f(x_i \mid \Pi, \Theta) = \prod_{i=1}^{N} \left[\sum_{j=1}^{K} \pi_{ij} \Phi(x_i \mid \Theta_j) \right] \quad (4)$$

This modeling has a fundamental problem. Since the observation x_i is considered to be independent given the pixel label, the spatial correlation between the neighboring pixels is not taken into account. In natural images, the neighboring pixels are highly correlated if they belong to same object. If the correlation is not used, the segmention can be very sensitive to noise, varying illumination and other environmental factors such as wind, rain or camera movements. MRF was introduced for segmentation in order to use this spatial information and has the following form:

$$p(\Pi) = Z^{-1} \exp\left\{ -\frac{1}{T} U(\Pi) \right\} \quad (5)$$

where, Z is a normalizing constant, T is a temperature constant set to 1 ($T = 1$), and $U(\Pi)$ is the smoothing prior. The posterior probability density function given by Bayes rules can be written as:

$$p(\Pi, \Theta \mid X) \propto p(X \mid \Pi, \Theta) p(\Pi) \quad (6)$$

By incorporating 4 and 5 into 6, the log-likelihood of 6 can be derived as:

$$\begin{aligned} L(\Pi, \Theta \mid X) &= \log p(\Pi, \Theta \mid X) \\ &= \sum_{i=1}^{N} \log \left\{ \sum_{j=1}^{K} \pi_{ij} \Phi(x_i \mid \Theta_j) \right\} - \log Z - \frac{1}{T} U(\Pi) \end{aligned} \quad (7)$$

Depending on the type of energy $U(\Pi)$ selected in Eq.(7), we can have different kinds of models. Different researchers have used different expressions for this energy function to successfully incorporate local information into the approach. But, this incorporation increases the complexity of the method and may not provide robustness against noise. Also, in order to maximize the log-likelihood function with respect to parameters Π and Θ, an iterative Expectation-Maximization (EM) algorithm needs to be applied. Due to the complexity of the log-likelihood function and the constraint in Eq.(2) to be satisfied, the M step of the EM algorithm cannot be directly applied to the prior distribution π_{ij}. Thus, the methods tend to become complex to solve the constrainted optimization problem.

3. THE PROPOSED METHOD

The proposed method is based on the fact that the energy function $U(\Pi)$ incorporates the spatial relationship among neighboring pixels and is a smoothing function that reduces the classification ambiguity between neighboring pixels. The method has been introduced keeping in mind that it should reduce the misclassification noise and in process, should not increase the computational complexity of the GMM.

In keeping with the above, we can refer to the following assumption. If the posterior probability of the i^{th} pixel for the j^{th} label is termed as z_{ij}, then the set $Z_j = \{z_{ij}; i = (1, 2, ..., N)\}$ for $j = (1, 2, ..., K)$ represents a posterior probability map which is smoothed using the energy function $U(\Pi)$ based on the neighboring relationship among the z_{ij} values. This assumption leads us to use image processing filters for smoothing this posterior probability map. Let, $\tilde{Z}_j$ represent the filtered posterior probability map after applying a filter to Z_j. If the elements of $\tilde{Z}_j$ are termed as $\tilde{z}_{ij}$, then we use the following approach to incorporate the spatial information into the smoothing prior $U(\Pi)$ as follows:

$$U(\Pi) = -\sum_{i=1}^{N} \sum_{j=1}^{K} \tilde{z}_{ij}^{(t)} \log \pi_{ij}^{(t+1)} \quad (8)$$

where, t indicates the iteration step. An important concern when applying a smoothing filter to Z_j is the edges where probability changes suddenly. This leads to application of an edge-preserving filter, which is discussed in detail in section 4. Considering a smoothed $\tilde{Z}_j$, the MRF distribution $p(\Pi)$ in Eq.(5) is given by:

$$p(\Pi) = Z^{-1} \exp\left\{ \frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{K} \tilde{z}_{ij}^{(t)} \log \pi_{ij}^{(t+1)} \right\} \quad (9)$$

Given the MRF distribution $p(\Pi)$, the log-likelihood function in Eq.(7) is written in the form:

$$\begin{aligned} L(\Pi, \Theta \mid X) = &\sum_{i=1}^{N} \log \left\{ \sum_{j=1}^{K} \pi_{ij} \Phi(x_i \mid \Theta_j) \right\} - \log Z \\ &+ \frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{K} \tilde{z}_{ij}^{(t)} \log \pi_{ij}^{(t+1)} \end{aligned} \quad (10)$$

Applying the complete data condition, maximization of $L(\Pi, \Theta \mid X)$ will lead to an increase in the value of the objective function $J(\Pi, \Theta \mid X)$ given by:

$$\begin{aligned} J(\Pi, \Theta \mid X) = &\sum_{i=1}^{N} \sum_{j=1}^{K} z_{ij}^{(t)} \left\{ \log \pi_{ij}^{(t+1)} + \log \Phi(x_i \mid \Theta_j^{(t+1)}) \right\} \\ &- \log Z + \frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{K} \tilde{z}_{ij}^{(t)} \log \pi_{ij}^{(t+1)} \end{aligned} \quad (11)$$

The conditional expectation values z_{ij} of the hidden variables can be computed as follows:

$$z_{ij}^{(t)} = \frac{\pi_{ij}^{(t)} \Phi(x_i \mid \Theta_j^{(t)})}{\sum_{k=1}^{K} \pi_{ik}^{(t)} \Phi(x_i \mid \Theta_k^{(t)})} \quad (12)$$

The next objective is to optimize the parameter set $\{\Pi, \Theta\}$ in order to maximize the objective function $J(\Pi, \Theta \mid X)$ in Eq.(11). For simplicity, Z and T in Eq.(11) are set equal to one ($Z = 1, T = 1$). From Eq.(11) and using Eq.(3), the objective function can be rewritten as:

$$\begin{aligned} &J(\Pi, \Theta \mid X) = \\ &\sum_{i=1}^{N} \sum_{j=1}^{K} z_{ij}^{(t)} \left\{ \log \pi_{ij}^{(t+1)} - \frac{D}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_j^{(t+1)}| \right\} \\ &+ \sum_{i=1}^{N} \sum_{j=1}^{K} z_{ij}^{(t)} \left\{ -\frac{1}{2} (x_i - \mu_j^{(t+1)})^T \Sigma_j^{-1(t+1)} (x_i - \mu_j^{(t+1)}) \right\} \\ &+ \sum_{i=1}^{N} \sum_{j=1}^{K} \tilde{z}_{ij}^{(t)} \log \pi_{ij}^{(t+1)} \end{aligned} \quad (13)$$

Table 1. Performance of proposed method with varying level of noise and varying spatial variance

Filter Sigma	Noise Sigma	Noisy MCR	MCR (Proposed)
3	0.03	9.66	0.51
	0.07	22.56	1.79
	0.1	27.48	3.06
5	0.03	9.66	1.11
	0.07	22.56	3.81
	0.1	27.48	8.04
9	0.03	9.66	2.56
	0.07	22.56	13.6
	0.1	27.48	21.59

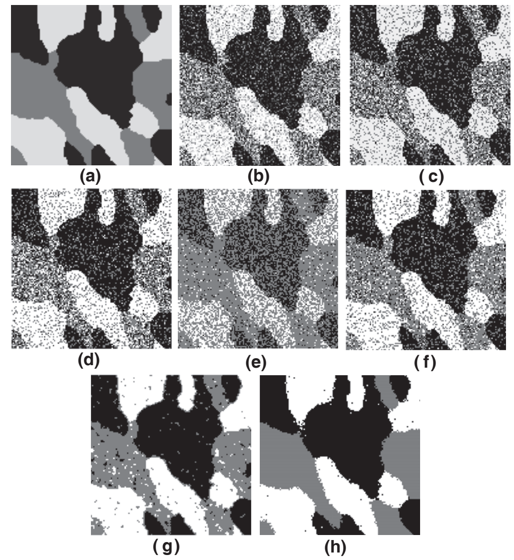

Fig. 1. Synthetic image segmentation, (a): original image, (b): image corrupted with noise, (c): K-means, (d): GMM, (e): SMM, (f): SVFMM, (g): FCM and (h): proposed

To maximize this function, the EM algorithm is applied where the derivative of $J(\Pi, \Theta \mid X)$ is taken with respect each parameter in the parameter set $\{\Pi, \Theta\}$ and equating it to zero. The solution to $\partial J/\partial \mu^{(t+1)_j} = 0$, $\partial J/\partial \Sigma_j^{-1(t+1)} = 0$ would provide the minimizer of μ_j and Σ_j respectively, at the $(t+1)$ step. It can be proven using simple vector differention, the minimizer values are:

$$\mu^{(t+1)_j} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} x_i}{\sum_{i=1}^{N} z_{ij}^{(t)}} \tag{14}$$

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} (x_i - \mu_j^{(t+1)})(x_i - \mu_j^{(t+1)})^T}{\sum_{i=1}^{N} z_{ij}^{(t)}} \tag{15}$$

For the prior distribution $\pi_{ij}^{(t+1)}$, the solution to $\partial J/\partial \pi_{ij}^{(t+1)} = 0$ must also satisfy the constraints in Eq.(2). To enforce the constraint, the Lagranges multiplier λ_i for each data point is used to get the following equation:

$$\frac{\partial}{\partial \pi_{ij}^{(t+1)}} \left[J - \sum_{i=1}^{N} \lambda_i \left(\sum_{j=1}^{K} \pi_{ij}^{(t+1)} - 1 \right) \right] \tag{16}$$

Eq.(16) can be solved using the constraint $\sum_{j=1}^{K} pi_{ij}^{(t+1)} = 1$ to yield the following solution:

$$pi_{ij}^{(t+1)} = \frac{z_{ij}^{(t)} + \tilde{z}_{ij}^{(t)}}{\sum_{k=1}^{K} (z_{ik}^{(t)} + \tilde{z}_{ik}^{(t)})} \tag{17}$$

Thus, using Eq.(14), (15) and (17), the optimum parameter values can be obtained that minimize J and hence, L.

4. BILATERAL FILTERING

In Sec.3, it was mentioned that the smoothed posterior probability map $\tilde{Z}_j$ is obtained using some image processing filter on the posterior probability map Z_j. A smoothing filter removes noise but at the same time blurs the image so that the edge information in the image is reduced. In segmentation, edges carry high importance and the borderline between two distinctly segmented regions is decided by how strong the edges are. Also, the edges in Z_j correspond to edges in the image because, in general, an edge signifies two clusters and hence, two different probabilities. This leads us to apply a filter that can preserve the edge information in high extent while smoothing the map (Note: In Z_j, an edge actually corresponds to a sudden change in probability). Bilateral filtering, in simple terms, is an *edge-preserving smoothing* filtering technique. Here, each pixel value (probability value in this case) is replaced by a weighted average of intensity values from neighboring pixels based on Gaussian distributions that are based on both the Euclidean distance and the range of intensity values of the neighboring pixels. Due to the combined distance based smoothing and intensity range based smoothing approach, the filter achieves the desired edge preservation.

When there is a non-edge region, the neighboring pixels have similar intensity and thus, bilateral filter acts as a standard smoothing filter that averages the noisy pixels with neighboring pixels. But, at the edges where there is a sudden change in intensity, part of the neighborhood have dark intensity and the rest are bright. In this case, due to a normalizing function, the center pixel value is replaced by the averaging values of the pixels in its vicinity. Thus, if the pixel belongs to dark region, its value will most likely be replaced by averaging the dark pixel values in its neighborhood. Similar reasoning applies for bright pixels. For mathematical basis of Bilateral filtering, the readers are referred to [12].

5. EXPERIMENTAL RESULTS

The proposed algorithm has been tested on the images from the Berkeley Image Segmentation dataset. The algorithm has been compared with K-means, Fuzzy C-Means Clustering (FCM) [4], standard GMM, Student's-t mixture model (SMM) [13] and Spatially Variant Finite Mixture Model (SVFMM) [14] algorithms which are some of the popular and leading methods for segmentation. The methods were run until convergence. The experimentation has been divided into two categories - (A) with a synthetic image for varying levels of noise and (B) with real world color images. All the methods were run on a PC with Intel Core 2 Duo CPU of 2 GHz with 2 GB of RAM. For synthetic image, in order to compare the results, misclassification ratio (MCR) has been used. MCR is given by the number of misclassified pixels divided by the total number of pixels. In the experiments, all the methods have been initialized with K means.

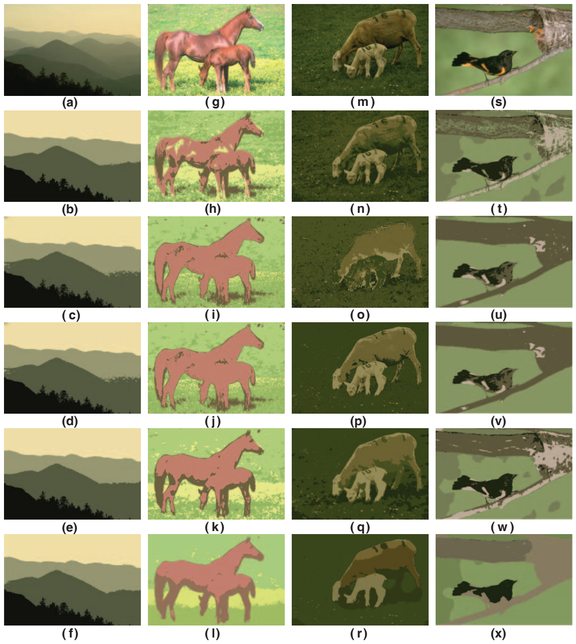

Fig. 2. Color image segmentation, (first row): original image, (second row): FCM, (third row): GMM, (fourth row): SMM, (fifth row): SVFMM, (sixth row): proposed

5.1. Segmentation of Synthetic Image

The algorithms were compared with a number of synthetic images with varying level of noise. In this work, results are shown with a single synthetic image for three levels of noise and effect of changing the spatial standard deviation value (sigma) of the Bilateral filter. The synthetic image has been corrupted with Gaussian noise with zero mean and varying variance value. One set of result is shown in Fig. 1 for a single noise level (0 mean, 0.1 variance) and for a constant spatial sigma 3. For varying level of noise, the MCR values for the proposed method are compared for varying spatial sigma values in Table 1.

From Fig. 1, it is visible that the performance of the proposed method is less affected by the noise. The parameters of the filter also controls the robustness of the method against varying noise level. The change in performance due to change in parameters and change in noise level can be observed from Table 1.

5.2. Segmentation of Real World Color Images

In this section, four real world color images are used from Berkeley dataset for comparing different methods. As a metric for comparison, Probabilistic Rand Index (PR Index) has been used. PR counts the fraction of pairs of pixels whose labelings are consistent between the computed segmentation and the ground truth, averaging across multiple ground truth segmentations to account for scale variation in human perception. The images are shown in Fig. 2 and the quantitative results are provided in Table 2.

From the figures, the effect of noise is noticeable. The figures 2(i), 2(k), 2(p), 2(q), 2(v) and 2(w) show how affected the segmentations are with the noisy pixels. The proposed method, on the other hand, is quite robust to this noise level and successfully segment the images into separate regions. The quantitative results are shown in Table 2. As can be seen, the proposed method has the highest PR values for the segmented images.

Table 2. Comparison of performance of proposed method with other methods for real-world color images

Images	FCM	GMM	SMM	SVFMM	Proposed
Moutains	0.889	0.888	0.889	0.887	**0.891**
Horse	0.750	0.782	0.810	0.777	**0.818**
Lamb	0.580	0.844	0.750	0.785	**0.856**
Bird	0.732	0.738	0.797	0.733	**0.808**

6. CONCLUSION

In this work, a new mixture model has been presented for image segmentation. The model uses simple bilateral filtering based MRF to include spatial relationship among neighboring pixels. Also, it has been kept fairly easy to manipulate the parameters of the technique and use the EM algorithm to compute the optimum values for the parameters of the mixture model. The future work would include searching for better methods to set the parameters of the filter automatically based on the image to be segmented.

7. REFERENCES

[1] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[2] P. F. Felzenswalb and D. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[3] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888905, 2000.

[4] C. Weiling, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognition*, vol. 40, no. 3, pp. 825838, 2007.

[5] S. Z. Li, "Markov random field modeling in image analysis," *SpringerVerlag*, 2009.

[6] Jain A. K., R. P. W. Duin, and J. C. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 437, 2000.

[7] D. M. Titterington, A. F. M. Smith, and U. E. Makov, "Statistical analysis of finite mixture distributions," *Hoboken, NJ: Wiley*, 1985.

[8] G. Celeux, F. Forbes, and N. Peyrard, "Em procedures using mean field-like approximations for markov model-based image segmentation," *Pattern Recognition*, vol. 36, no. 1, pp. 131–144, 2003.

[9] F. Forbes and N. Peyrard, "Hidden markov random field model selection criteria based on mean field-like approximations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 10891101, 2003.

[10] C. Nikou, N. Galatsanos, and A. Likas, "A class-adaptive spatially variant mixture model for image segmentation," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 11211130, 2007.

[11] M. Robinson, M. Azimi-Sadjadi, and J. Salazar, "A temporally adaptive classifier for multispectral magery," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pp. 159165, 2004.

[12] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "Bilateral filtering: Theory and applications," *Foundations and Trends in Computer Graphics and Vision*, vol. 4, no. 1, pp. 1–73, 2009.

[13] G. Sfikas and C. Nikou abd N. Galatsanos, "Robust image segmentation with mixtures of students t-distributions," *IEEE International Conference on Image Processing*, vol. 1, pp. 27327, 2007.

[14] K. Blekas, A. Likas, N. P. Galatsanos, and I. E. Lagaris, "A spatially constrained mixture model for image segmentation," *IEEE Transactions on Neural Networks*, vol. 16, no. 2, pp. 494–498, 2005.