Accurate image segmentation using Gaussian mixture model with saliency map

Hui Bi, Hui Tang, Guanyu Yang, Huazhong Shu, Jean-Louis Dillenseger

Noname manuscript No.
(will be inserted by the editor)

Accurate Image Segmentation Using Gaussian Mixture Model with Saliency Map

Hui Bi · Hui Tang · Guanyu Yang · Huazhong Shu · Jean-Louis Dillenseger

Received: 24 February 2017 / Accepted: 8 December 2017

Abstract Gaussian mixture model (GMM) is a flexible tool for image segmentation and image classification. However, one main limitation of GMM is that it doesn't consider spatial information. Some authors introduced global spatial information from neighbor pixels into GMM without taking the image content into account. The technique of saliency map, which is based on the human visual system, enhances the image regions with high perceptive information. In this paper, we propose a new model, which incorporates the image content-based spatial information extracted from saliency map into the conventional GMM. The proposed method has several advantages: it is easy to implement into the Expectation Maximization algorithm for parameters estimation and therefore there is only little impact in computational cost. Experimental results performed on the public *Berkeley* database show that the proposed method outperforms the state-of-art methods in terms of accuracy and computational time.

Keywords Image segmentation · Gaussian mixture model · Spatial information · Saliency map · Object recognition

1 Introduction

IImage segmentation plays an important role in artificial intelligence and image understanding [1, 2]. Over the past decades, works on automatic image segmentation has been a growing interest. Various categories of models for image segmen-

H. Bi and H. Tang and GY Yang and HZ Shu
Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing, China;
Key Laboratory of Computer Network and Information Integration, Ministry of Education, Nanjing, China;

J-L. Dillenseger
NSERM U1099, 35000 Rennes, France;
Laboratoire Traitement du Signal et de lImage, Universit de Rennes I, 35000 Rennes, France;

HZ Shu and J-L. Dillenseger
Centre de Recherche en Information Biomdicale sino-franais (CRIBs), Nanjing, China

Table 1 Summary.of methods including spatial information in GMM.

Principle	References	Advantages	Disadvantages
Markov Random Field	[16–20]	High segmentation accuracy, adaptation to image content	High computational cost
Mean Template	[21–23]	Simple to implement, fast computation, robust and effective	No adaptation to image content

tation models has been explored, such as edge detection, texture analysis, or finite mixture model [3].

Among finite mixture models, Gaussian mixture model (GMM) is the most common tool used for image segmentation [4, 5] segmentation or video object segmentation [6–10]. The Expectation Maximization (EM) algorithm is often used to estimate the parameters of the distribution [11–15].

However, as any finite mixture models, GMM does not consider image spatial information. In fact, the classical GMM considers each pixel as independent but in an image the objects of interest are composed by connected pixels which share some common statistical properties: values, colors, textures, etc. Many methods have been proposed to incorporate the spatial information in order to improve the conventional GMM [16, 17]. A common way to handle neighboring pixels dependencies is the use of Markov random field (MRF) [18]. SSo, the incorporation of spatial information in mixtures model based on MRF has been proposed for image segmentation [16, 19, 20, 17]. But these methods suffer from two drawbacks: (1) in the parameters learning step, the model parameters cannot be estimated directly in the Maximization step (M-step) of the EM algorithm and (2) the use of MRF is computationally expensive. Although MRF-based GMM show excellent segmentation results, the very high computational cost limits its use in practical application. Another approach consists to incorporate directly some local spatial information using a mean template (GMM-MT) [21]. GMM-MT has later been extended by applying either a weighted arithmetic or a weighted geometric mean template to the conditional and the prior probability, called ACAP, ACGP, GCGP, and GCAP [22, 23]. These four models are robust to noise and fast to implement. However, these weights are generally equally assigned to the neighbor pixels without any content. A summary of these approaches is listed in 1.

Recently, the visual saliency becomes a popular topic for object recognition. This class of methods is based on modeling the visual attention system inspired by the neuronal architecture and the behavior of the primate early visual system. When the goal of an application is the object recognition in an image, a visual saliency map is constructed by the combination of multiscale low-level image features, such as intensities, colors, and orientations. These features try to identify the most informative parts on an image which are candidates to belong to an object [24, 25]. Rensink used the saliency map to detect the region of interest in an image and introduced the notion of proto-objects in [26–28]. Itti and Koch proposed a framework for saliency detection that breaks down the complex problem of image understanding by a rapid and computationally efficient selection of conspicuous locations [29, 30]. Then, this group extended the saliency model to object recognition tasks [31]. However, the image features-based saliency map extraction is computationally expensive. To overcome this shortcoming, Hou [32] proposed a

simple way to extract a saliency map using the spectral residual in the spectral domain. Essentially a saliency map reflects the visual importance of each pixel in one image. Therefore, it points out the most noteworthy regions and introduces also rough content-based information.

In this paper we propose to use a saliency map to incorporate context-based spatial information into the conventional GMM for image segmentation. Our model, known as GMM with Spatial Information extracted from Saliency Map (GMM-SMSI), is divided into two main steps. Firstly, a saliency map detection is obtained by means of the image spectral residual. Secondly, the saliency map is incorporated as spatial information into the conventional GMM. This two steps approach allowed us to adapt the neighboring template of GMM-MT according to the image content. The proposed model should improve the classical GMM scheme because (1) the saliency map can directly incorporate some spatial information by means of some specific weights assigned to neighbor pixels of the current pixel; (2) it is easy to implement since the saliency detection is an independent step from GMM.

The paper is organized as follows. Section 2 describes the proposed method in detail. In Section 3 we present and discus experimental results. Finally section 4 concludes this paper.

2 GMM with Saliency Map (GMM-SMSI)

2.1 Saliency Map

Based on the human visual system, the concept of saliency map has been developed for image understanding and object recognition. The saliency map reflects the regions of an image, which can present an interest in the sense of visual perception. It highlights the pixels, which can potentially contain information to be used in a more complex image classification scheme. In computer vision, visual attention usually focuses on unexpected features in an image. The basic principle of saliency is to suppress the response of frequently occurring features and to keep abnormal features. Researchers made use of the similarities that occur in training images to explore redundant information and so to detect the saliency map. However, this leads to heavy computation cost. To reduce the computational complexity, it is worth to explore solutions where only one individual image can be used to realize the saliency detection [24–31]. Some researchers tried to extract saliency maps by making use of information in the spectral domain instead of the spatial domain [32–34]. These methods are based on the fact that each image share some statistical redundant average information and can be differentiated by some statistical singularities.

In the spectral domain, the statistical redundant average information can be estimated by the average spectrum $A(f)$ of an image which suggests a local linearity [32]. For an individual image, $A(f)$ can be approximated by filtering the image log spectrum $L(f)$ by a local average filter $h_n(f)$:

$$A(f) = h_n(f) * L(f) \tag{1}$$

where $h_n(f)$ is an $n \times n$ mean convolution filter defined by:

$$h_n(f) = \frac{1}{n^2} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ & & & \\ 1 & 1 & \ddots & 1 \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

The statistical singularities of each image can be reflected by a spectrum residual $R(f)$ given by:

$$R(f) = L(f) - A(f) \tag{2}$$

The saliency map $S(x)$ is then calculated based on the residual spectrum as:

$$S(x) = G(x) * F^{-1} \left\{ \exp \left[R(f) + P(f) \right] \right\}^2 \tag{3}$$

where $G(x)$ denotes a Gaussian filter to smooth the resulting map, F^{-1} is the inverse Fourier transform and $P(f)$ denotes the phase spectrum of the image which is preserved during the processing.

Fig. 1 shows some examples of image saliency maps. The first column shows the original images. The second column shows the log spectrum of the image (blue solid line) and the spectrum residual (red solid line). It has to be noted that the scales are not the same between these curves. The third column shows the corresponding saliency map. The saliency map is an explicit representation of proto-objects. Most of the authors uses a simple threshold in the saliency map to detect the proto-objects. As shown in Fig. 2, this method achieves good results in the images saliency map extraction.

In our approach, the saliency map is used to assess the regional context around a specific pixel, i.e., the influence of the neighborhood. In this way, each pixel is no longer considered as an individual but influenced by the neighborhood information.

2.2 Saliency Weighted GMM

GMM is a probabilistic model which represents a distribution by a simple linear combination of Gaussian densities. GMM can be used to cluster N pixels into L class labels [3]. Consider the following symbols: $i \in 1, 2, , N$ denotes an image pixel index, x_i is the ith pixel in image, $j \in 1, 2, , L$ represents the class label index. In the conventional GMM, the conditional probability that x_i belongs to class j is given by:

$$\Phi(x_i|\theta_j) = \frac{1}{2\pi|\Sigma_j|} \exp \left[-\frac{1}{2}(x_i - \mu_j)\Sigma_j^{-1}(x_i - \mu_j) \right] \tag{4}$$

where $\theta_j = \{\mu_j, \Sigma_j\}$ denotes the mean and the variance of the jth Gaussian distribution.

In a conventional GMM, the pixel value distribution can be described by the following equation:

$$f(x_i|\Pi, \Theta) = \sum_{j=1}^{L} \pi_j \Phi(x_i|\theta_j) \tag{5}$$

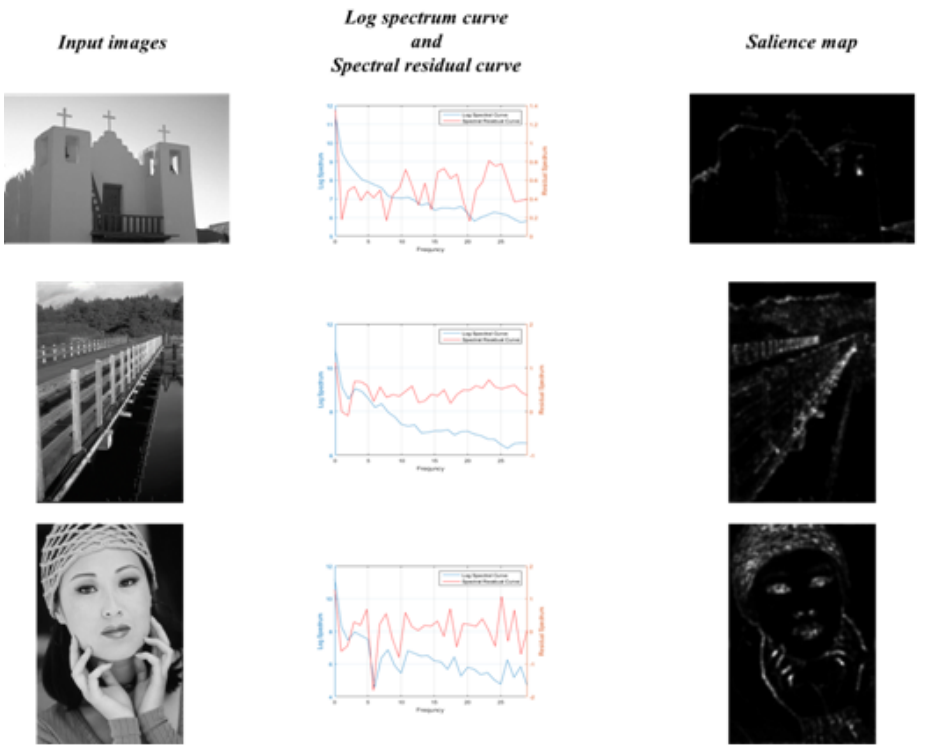

Fig. 1 Saliency map construction. The first column shows the original images. The second column shows the log spectrum (blue solid line) and the spectrum residual (red solid line) of the corresponding image. The saliency maps are on the third column.

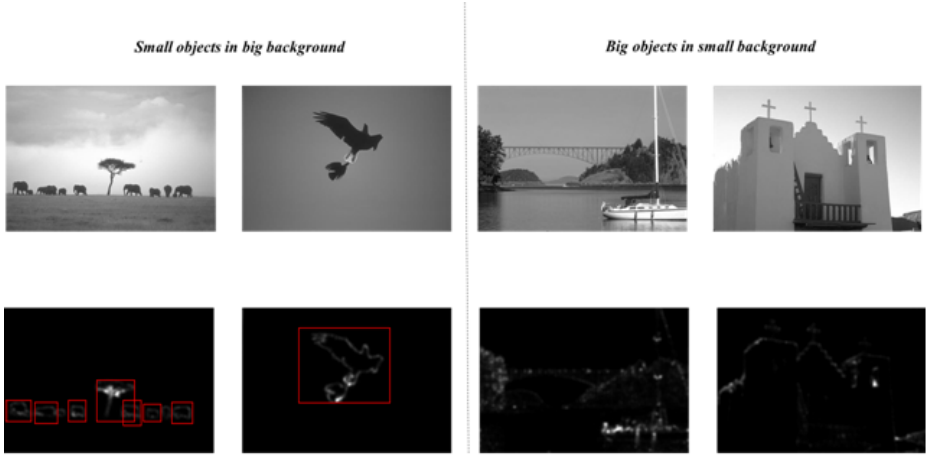

Fig. 2 . Proto-objects extraction based on saliency map.

where $\Pi = pi_1, pi_2, , pi_L$ denotes the set of prior probabilities (also called mixture component weights), and $\Theta = \theta_1, \theta_2, , \theta_L$ is the set of parameters of all Gaussian distributions.

The spatial information can be introduced in the GMM as a weighted neighborhood template for computing the conditional probability of x_i by its neighborhood probabilities [21,23]. In our case, we use directly the saliency map $S(x)$ to assign the proper neighborhood weights. The Saliency weighted GMM is then:

$$f(y_i|\Psi) = \sum_{j=1}^{L} \pi_{ij} \left[\sum_{m \in N_i} \frac{S(x_m)}{R_i} p(y_m|\theta_j) \right] \tag{6}$$

where π_{ij} denotes the probability that the pixel x_i belongs to class j, π_{ij} satisfies the constraints $\pi_{ij} \geq 0$ and $\sum_{j=1}^{L} \pi_{ij} = 1$; N_i is the neighborhood of the pixel x_i; R_i is the sum of the saliency map values inside N_i; Ψ denotes the parameters set containing all the parameters: $\Psi = \{\pi_{11}, \pi_{12}, \cdots, \pi_{1L}, \pi_{21}, \pi_{22}, \cdots, \pi_{2L}, \pi_{N1}, \pi_{N2}, \cdots, \pi_{NL}, \theta_1, \theta_2, \cdots, \theta_L\}$ and $S(x_m)$ is the saliency map value at location x_m.

We then apply the EM algorithm for the parameter estimation in our model. According to [3], the complete-data log likelihood function is calculated as follows:

$$Q = \sum_{i=1}^{N} \sum_{j=1}^{L} \gamma_{ij} \left[\sum_{m \in N_i} \frac{S(x_m)}{R_i} p(y_m|\theta_j) + \log \pi_{ij} \right] \tag{7}$$

In the Expectation step (E-step) the posterior probability can be calculated as follows:

$$\gamma_{ij}^{(t)} = \frac{\pi_{ij}^{(t)} \sum_{m \in N_i} \frac{S(x_m)}{R_i} p(y_m|\theta_j^{(t)})}{\sum_{h=1}^{L} \pi_{ih}^{(t)} \sum_{m \in N_i} \frac{S(x_m)}{R_i} p(y_m|\theta_j^{(t)})} \tag{8}$$

In the maximization step (M-step), the mean and covariance are computed as follows:

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{N} \sum_{m \in N_i} \gamma_{ij}^{(t)} \frac{S(x_m)}{R_i} x_m}{\sum_{i=1}^{N} \gamma_{ij}^{(t)}} \tag{9}$$

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{N} \sum_{m \in N_i} \gamma_{ij}^{(t)} \frac{S(x_m)}{R_i} (x_m - \mu_j^{(t)})(x_m - \mu_j^{(t)})^T}{\sum_{i=1}^{N} \gamma_{ij}^{(t)}} \tag{10}$$

And the prior probability is given by:

$$\pi_{ij}^{(t+1)} = \frac{\sum_{m \in N_i} S(x_m) \gamma_{mj}^{(t)}}{\sum_{h=1}^{L} \sum_{m \in N_i} S(x_m) \gamma_{mh}^{(t)}} \tag{11}$$

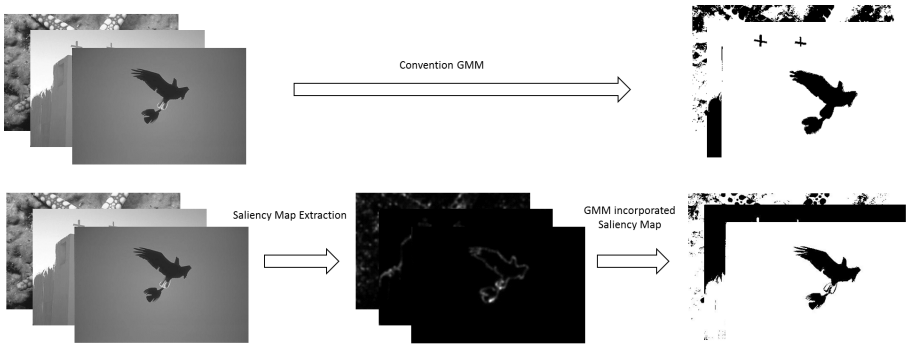

Fig. 3 Flow chart of the conventional GMM and GMM-SMSI

2.3 Flow Chart of GMM-SMSI

The flow chart of the proposed model is described as follows:

Step 1: Salience Map Extraction.

1. The image is converted in the spectral domain using FFT. This gives the amplitude spectrum $F(f)$ and the phase spectrum $P(f)$ of the image.
2. The log spectrum representation $L(f)$ is given by the logarithm of $F(f)$.
3. The estimation of the average spectrum $A(f)$ is given by using (1).
4. The calculation of the residual value $R(f)$ is given by using (2).
5. The generation of the saliency map $S(x)$ is given by using (3).

Step 2: GMM incorporating the saliency map as spatial information.

1. The k-means algorithm is first used to initialize the parameters set $\Psi^{(0)}$
2. Using the saliency map $S(x)$ the EM algorithm (Eqs. (8) - (11)) is applied for the parameters estimation until convergence. At the end, we get the parameters set $\Psi(c)$.
3. The image pixels are then classified (labeled) based on the highest posterior probability.

A brief example of the flow chart of the conventional GMM and GMM-SMSI are shown in Fig. 3. Compared to the conventional GMM, our algorithm does not segment directly the data based on the posterior maximum value but extracts a saliency map firstly and incorporates it as weight to perform the GMM algorithm.

3 Experiments

In this section, experimental results of GMM-SMSI are compared with some classical methods such as Spatial Variant Finite Mixture Model (SVFMM) [20], Fuzzy Local Information C-Means (FLICM) [35], Hidden Markov Random Field with Fuzzy C-Means (HMRF-FCM) [18], and the Mean Template GMM variant in which an arithmetic mean template is incorporated in both the Conditional and Prior probabilities (ACAP) [23]. Our experiments have been performed on MATLAB R2013a, and are run on an Intel i5 Core 2.8GHz CPU with 12.0GB RAM. We

experimentally evaluated these methods on a set of real images from the *Berkeley* image dataset [36].

The segmentation performance of these methods was evaluated using Probabilistic Rand (PR) index values [37]. It has been shown that the PR index takes values between 0 and 1. A PR value close to 1 indicates a better segmentation while close to 0 indicates a worse one. In order to analyze the behavior of the methods, we first performed the experiments on four classes of image content: a tiny object in a large background region, a large object in a small background region, buildings and human face images. Then we globally compared the overall performance of the several methods using the average PR and the average computation time when the methods were applied on all the *Berkeley* image dataset.

3.1 Tiny Object in large background region

In the first experiment, we chose an image (481×321) with 2 birds in the sky in order to show the ability to segment tiny objects in a large background region (Fig. 4 (a)). The goal was to segment the image into two classes: the objects with two birds and the sky. Comparing the results of the several methods, we noticed that for SVFMM (Fig. 4 (b), PR = 0.9835), there was a large misclassification of the sky and also of the region between the birds. FLICM (Fig. 4 (c), PR = 0.9834) showed a similar misclassification of the sky; however, the 2 birds were now disconnected. In HMRF-FCM (Fig. 4 (d), PR = 0.9853), the sky misclassification was smaller than SVFMM and FLICM, however, the two birds were difficult to separate. The accuracy of the segmentation for ACAP (Fig. 4 (e), PR = 0.9855) was better than the other three methods because there was a good classification of sky; however the two birds were not separated. Our method, GMM-SMSI, (Fig. 4 (f), PR = 0.9864) was able to distinguish the 2 birds. Compared to the other methods, the wings of the little bird (Green Square) showed also more details. Furthermore, our algorithm obtained the highest PR index value.

3.2 Large Object in small background region

In the second experiment, we chose an image (481×321) with a relatively large star-fish in the seabed (Fig. 5 (a)). We also tried to segment the image into two classes: the background and the star-fish. As shown in Fig. 5 (b), the segmentation for SVFMM (PR = 0.6835) was not able to distinguish the upper part of the starfish from the background. It can be seen in Fig. 5 (c) that FLICM (PR = 0. 6834) had a similar behavior compared with SVFMM. Fig. 5 (d) shows that HMRF-FCM (PR = 0.7853) achieved good background and object cluster separation whereas it was not suitable to segment an object with lot of details. The segmentation for ACAP (PR = 0.6855) was better than SVFMM, FLICM and HMCR-FCM, as shown in Fig. 5 (e); the starfish was clearly separated from the background. There were still some pixels misclassifications at the edge of the starfish. It can be noticed (Fig. 5 (f)) that our algorithm (PR = 0.6986) separated clearly the starfish and the background. The details of the starfish can be better distinguished. Furthermore, our algorithm obtained the highest PR index values.

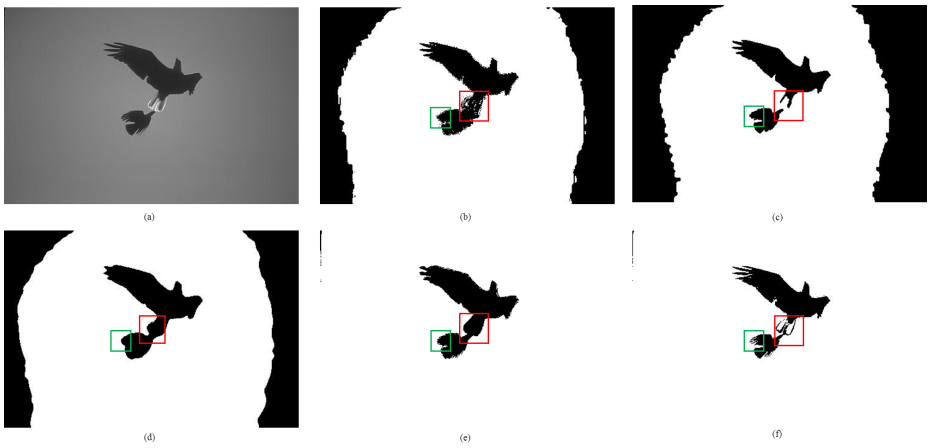

Fig. 4 Tiny object in a large background region segmentation results. (a) Original image. (b) SVFMM, PR = 0.9835. (c) FLICM, PR = 0.9834. (d) HMCR-FCM, PR = 0.9853. (e) ACAP, PR = 0.9855. (f) GMM-SMSI, PR = 0.9864.

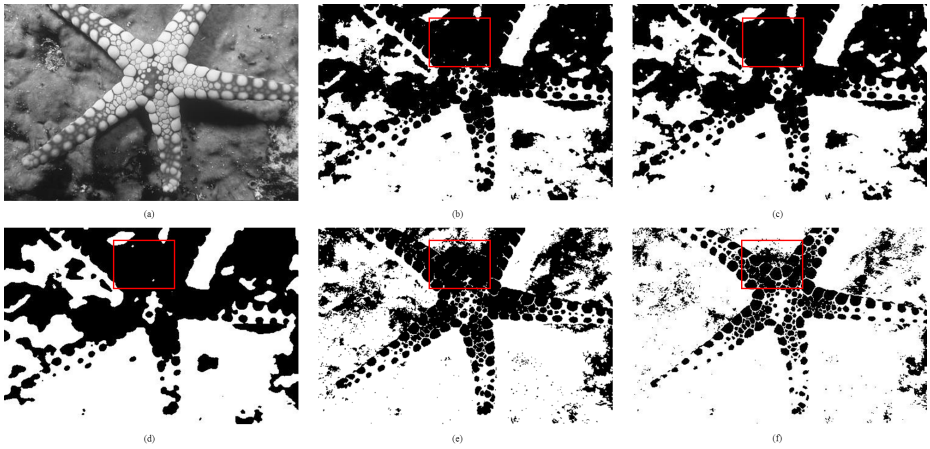

Fig. 5 Large object in a small background region segmentation results. (a) Original image. (b) SVFMM, PR = 0.6835. (c) FLICM, PR = 0.6834. (d) HMCR-FCM, PR = 0.7853. (e) ACAP, PR = 0.6855. (f) GMM-SMSI, PR = 0.6986.

3.3 Building

In the third experiment, we tried to segment an image of a building (481×321) into two classes: the church and the background (Fig. 6 (a)). It can be seen in Fig. 6 (b - d) that for SVFMM (PR = 0. 7204), FLICM (PR = 0. 8977) and HMCR-FCM (PR = 0. 8379), the top right corner of the background was misclassified as church region. The misclassification of SVFMM was much larger than FLCM and HMCR-FCM. The segmentation accuracy of ACAP (PR = 0.8599) was better, the sky and the church were clearly separated, as Fig. 6 (e) shown. The segmentation of our method (PR = 0. 8362) showed more details in the church, such as stairs, the left window and the door (Fig. 6 (f)). It was more in phase with human vision.

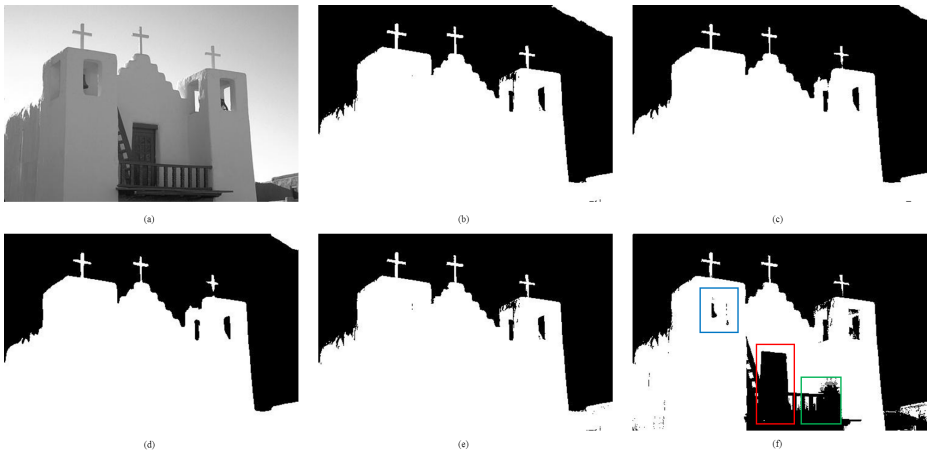

Fig. 6 Building image segmentation results. (a) Original image. (b) SVFMM, PR = 0.7204. (c) FLICM, PR = 0.8977. (d) HMCR-FCM, PR = 0.8379. (e) ACAP, PR = 0.8599. (f) GMM-SMSI, PR = 0.8362.

However, it has to be noted that the PR index value of our method was lower than FLICM, HMRF-FCM and ACAP since the door, the stairs, the window and even some shadow were classified as background in our method and as object in the ground truth. However, our algorithm showed more details and so offered more information for image understanding.

3.4 Human Face

We also performed our evaluation on a human face image (481 × 321) as shown in Fig. 7 (a). The goal was to segment the image into two classes. In Fig. 7 (b), the information about human face can be detected through SVFMM (PR = 0. 7204) whereas it contained only a part of the eyebrows rather than the whole ones. As shown in Fig. 7 (c), FLICM (PR = 0. 8977) provided correct facial information, however, the textures of the clothes were not clear. HMCR-FCM (PR = 0. 8379) achieved a better clustering than SVFMM and FLICM but with lot of lost information, as shown in Fig. 7 (d). Similar to HMRF-FCM, ACAP (PR = 0. 8599) also achieved better clustering (Fig. 7 (e)). On Fig. 7 (f), it can be seen that our algorithm (PR = 0.8362) showed more details of the human face, such as the full eyebrow information. It was also true when considering the texture of the clothes. It can also be noticed that the PR index value of our method was lower than FLICM, HMRF-FCM and ACAP. Actually, these methods proposed a better clustering but with details loss. However, our algorithm showed more details to offer more information for face recognition and image understanding.

3.5 Global segmentation performance

In this subsection some objective ways to evaluate SVFMM, FLICM, HMCR-FCM, ACAP and GMM-SMSI are proposed. Table 2 presents the PR index values

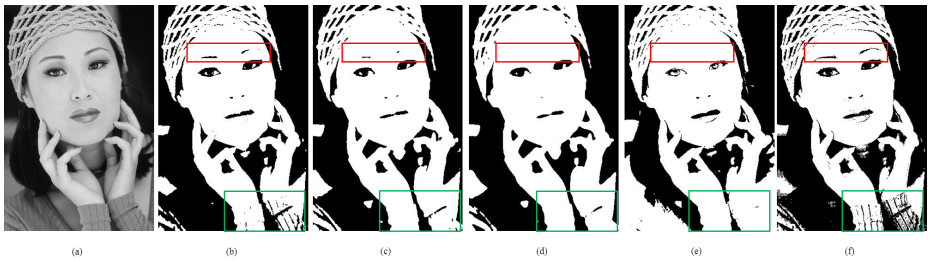

Fig. 7 Face image segmentation results. (a) Original image. (b) SVFMM, PR = 0.7204. (c) FLICM, PR = 0.8977. (d) HMCR-FCM, PR = 0.8379. (e) ACAP, PR = 0.8599. (f) GMM-SMSI, PR = 0.8362.

obtained on a sample of 9 different images. These images varied in terms of number of classes to be estimated. Globally our method obtained almost the best PR index (bold numbers) for these examples. This performance is confirmed by the mean of the PR indexes contained obtained by the several methods on all the images of the *Berkeley* image dataset. GMM-SMSI obtained the highest mean PR index (nearly 12.55% higher than SVFMM) and can so be considered to be globally the most accurate algorithm between the evaluated methods. Fig. 8 shows the boxplot of the PR indexes obtained by each method on the whole *Berkeley* image dataset. It confirms that GMM-SMSI had the highest median value but also the smallest interquartile range which indicates a higher robustness.

3.6 Computation time

In this subsection, we evaluate the computation time of SVFMM, FLICM, HMCR-FCM, ACAP and GMM-SMSI. Table 2 also presents the average computation time of each method when applying to the whole image set. GMM-SMSI took the lowest computation time that was nearly 14.67% of this of HMCR-FCM. The result can be explained by several facts: the spatial information is computed only once, so no further spatial research is needed; the spatial information is integrated explicitly in the EM scheme; and the spatial information helped the EM algorithm to converge faster. Based on the experiments, it appears that the proposed GMM-SMSI algorithm brought some benefits in aspects of accuracy, time-cost and the capability to display more detailed information.

4 Conclusion

In this paper, we have proposed a new algorithm, the Gaussian Mixture Model with Saliency Map as Spatial Information based on classical Gaussian Mixture Model. The saliency map helped to incorporate image content-based spatial information into the GMM. The conditional probability of an image pixel was replaced by the computation of the probabilities in its immediate neighborhood weighted by the image saliency map information. Saliency map assigned the proper weights to pixels neighborhood to enhance the role of significant pixels. Since the saliency map extraction was independent of GMM, it made the proposed model simple to

Table 2 Comparison between the different methods applied on the *Berkeley* image dataset, PR Index and mean computation time.

Image #	Class	SVFMM [20]	FLICM [35]	HMCR-FCM [18]	ACAP [23]	GMM-SMSI
78019	7	0.7790	0.6580	0.8308	0.8006	**0.8091**
106025	4	0.6347	0.8116	0.7988	0.8205	**0.8524**
253036	4	0.6442	0.8838	0.8690	0.9133	**0.9617**
24063	4	0.7204	**0.8977**	0.8372	0.8599	0.8362
61086	5	0.7143	0.6873	0.7299	0.7344	**0.8034**
22090	4	0.7752	0.7675	0.7777	0.8004	**0.8026**
302003	3	0.7170	0.7172	0.7169	0.7179	**0.7816**
Mean PR value on all the dataset Images		0.7294	0.7756	0.7923	0.8043	**0.8245**
Mean computation time (seconds)		28.42	55.16	82.91	13.62	**12.16**

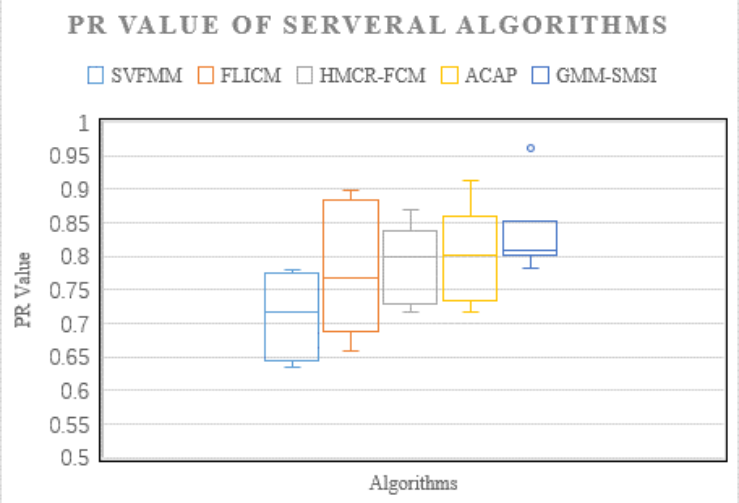

Fig. 8 Boxplot of the PR for the several algorithms applied on the *Berkeley* image dataset.

implement. In addition, the parameters of GMM-SMSI can be easily estimated by Expectation Maximum (EM) algorithm. In experiments performed on the public *Berkeley* database, we have demonstrated that the proposed GMM-SMSI method outperformed the state-of-art methods in aspects of both classification accuracy and computation time. Moreover, these experiments indicated that our method can detect more objects details in an image. In summary, the proposed GMM-SMSI is an accurate, robust and fast algorithm which can be easily implemented and has a good execution time performance.

Acknowledgements This work was supported by the by the National Natural Science Foundation of China under Grants 61271312, 61201344, 61401085, 31571001, and 81530060, and by Natural Science Foundation of Jiangsu Province under Grant BK2012329, BK2012743, BK20150647, DZXX-031, BY2014127-11, by the 333 project under Grant BRA2015288 and by the Qing Lan Project.

References

1. M. Sonka, V. Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*. Cengage Learning, 2014.
2. G. McLachlan and D. Peel, *Finite mixture models*. John Wiley & Sons, 2004.
3. C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.
4. N. Bouguila, "Count data modeling and classification using finite mixtures of distributions," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 186–198, 2011.
5. S. E. Yuksel, J. N. Wilson, and P. D. Gader, "Twenty years of mixture of experts," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 8, pp. 1177–1193, 2012.
6. T. Bouwmans, F. El Baf, and B. Vachon, "Background modeling using mixture of Gaussians for foreground detection-a survey," *Recent Patents on Computer Science*, vol. 1, no. 3, pp. 219–237, 2008.
7. F. El Baf, T. Bouwmans, and B. Vachon, "Type-2 fuzzy mixture of Gaussians model: application to background modeling," in *International Symposium on Visual Computing*. Springer, 2008, pp. 772–781.
8. M. S. Allili, N. Bouguila, and D. Ziou, "A robust video foreground segmentation by using generalized Gaussian mixture modeling," in *Computer and Robot Vision, 2007. CRV'07. Fourth Canadian Conference on*. IEEE, 2007, pp. 503–509.
9. ——, "Finite generalized Gaussian mixture modeling and applications to image and video foreground segmentation," in *Computer and Robot Vision, 2007. CRV'07. Fourth Canadian Conference on*. IEEE, 2007, pp. 183–190.
10. M. Shah, J. Deng, and B. Woodford, "Illumination invariant background model using mixture of Gaussians and SURF features," in *Asian Conference on Computer Vision*. Springer, 2012, pp. 308–314.
11. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B (methodological)*, vol. 39, pp. 1–38, 1977.
12. J. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," International Computer Science Institute, Tech. Rep., 1998.
13. T. Denœux, "Maximum likelihood estimation from fuzzy data using the EM algorithm," *Fuzzy sets and systems*, vol. 183, no. 1, pp. 72–91, 2011.
14. G. McLachlan and T. Krishnan, *The EM algorithm and extensions*. John Wiley & Sons, 2007, vol. 382.
15. M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, 2002.
16. 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, Mar. 2005.
17. T. M. Nguyen and Q. Wu, "Gaussian-mixture-model-based spatial neighborhood relationships for pixel labeling problem," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 1, pp. 193–202, 2012.
18. S. P. Chatzis and T. A. Varvarigou, "A fuzzy clustering approach toward hidden Markov random field models for enhanced spatially constrained image segmentation," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 5, pp. 1351–1361, Oct. 2008.
19. A. Diplaros, N. Vlassis, and T. Gevers, "A spatially constrained generative model and an EM algorithm for image segmentation," *IEEE Transactions on Neural Networks*, vol. 18, no. 3, pp. 798–808, May 2007.
20. S. Sanjay-Gopal and T. J. Hebert, "Bayesian pixel classification using spatially variant finite mixtures and the generalized em algorithm," *IEEE Transactions on Image Processing*, vol. 7, no. 7, pp. 1014–1028, Jul. 1998.
21. H. Tang, J.-L. Dillenseger, X. D. Bao, and L. M. Luo, "A vectorial image soft segmentation method based on neighborhood weighted gaussian mixture model." *Computerized Medical Imaging and Graphics*, vol. 33, no. 8, pp. 644–650, Dec 2009.
22. H. Zhang, Q. M. J. Wu, and T. M. Nguyen, "Image segmentation by a robust modified gaussian mixture model," in *Proc. Speech and Signal Processing 2013 IEEE Int. Conf. Acoustics*, May 2013, pp. 1478–1482.

23. ——, "Incorporating mean template into finite mixture model for image segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 2, pp. 328–335, Feb. 2013.
24. A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive psychology*, vol. 12, no. 1, pp. 97–136, 1980.
25. M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.
26. R. A. Rensink and J. T. Enns, "Preemption effects in visual search: evidence for low-level grouping." *Psychological review*, vol. 102, no. 1, p. 101, 1995.
27. R. A. Rensink, J. K. O'Regan, and J. J. Clark, "To see or not to see: The need for attention to perceive changes in scenes," *Psychological science*, vol. 8, no. 5, pp. 368–373, 1997.
28. R. A. Rensink, "Seeing, sensing, and scrutinizing," *Vision research*, vol. 40, no. 10, pp. 1469–1487, 2000.
29. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
30. A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013.
31. D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch, "Attentional selection for object recognitiona gentle way," in *Biologically motivated computer vision*. Springer, 2002, pp. 251–267.
32. X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.
33. D. L. Ruderman, "The statistics of natural images," *Network: computation in neural systems*, vol. 5, no. 4, pp. 517–548, 1994.
34. A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *Journal of mathematical imaging and vision*, vol. 18, no. 1, pp. 17–33, 2003.
35. S. Krinidis and V. Chatzis, "A robust fuzzy local information c-means clustering algorithm," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1328–1337, 2010.
36. D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2. IEEE, 2001, pp. 416–423.
37. R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 929–944, 2007.