Pattern Analysis and Applications (2018) 21:869–878
https://doi.org/10.1007/s10044-017-0672-1

INDUSTRIAL AND COMMERCIAL APPLICATION

Accurate image segmentation using Gaussian mixture model with saliency map

Hui Bi[1,2] · Hui Tang[1,2] · Guanyu Yang[1,2] · Huazhong Shu[1,2,3] · Jean-Louis Dillenseger[3,4,5]

Received: 24 February 2017 / Accepted: 8 December 2017 / Published online: 27 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 does not 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-the-art methods in terms of accuracy and computational time.

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

1 Introduction

Image segmentation plays an important role in artificial intelligence and image understanding [1, 2]. Over the past decades, works on automatic image segmentation have been a growing interest. Various categories of models for image segmentation models have 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] 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]. So, the incorporation of spatial information in mixtures model based on MRF has been proposed for image segmentation [16, 17, 19, 20]. But these methods

✉ Hui Bi
bihui@seu.edu.cn

Hui Tang
corinna@seu.edu.cn

Guanyu Yang
yang.list@seu.edu.cn

Huazhong Shu
shu.list@seu.edu.cn

Jean-Louis Dillenseger
jean-louis.dillenseger@univ-rennes1.fr

1 Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing, China

2 Key Laboratory of Computer Network and Information Integration, Ministry of Education, Nanjing, China

3 Centre de Recherche en Information Biomédicale sino-français (CRIBs), Nanjing, China

4 INSERM U1099, 35000 Rennes, France

5 Laboratoire Traitement du Signal et de l'Image, Université de Rennes I, 35000 Rennes, France

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 shows excellent segmentation results, the very high computational cost limits its use in practical applications. Another approach incorporates local spatial information directly by 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 Table 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-step 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 and (2) it is easy to implement since the saliency detection is an independent step from GMM.

The paper is organized as follows: Sect. 2 describes the proposed method in detail. In Sect. 3 we present and discus experimental results. Finally, Sect. 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 exploring the 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 shares 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 $\mathcal{A}(f)$ of an image which suggests a local linearity [32]. For an

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

individual image, $\mathcal{A}(f)$ can be approximated by filtering the image log spectrum $\mathcal{L}(f)$ by a local average filter $h_n(f)$:

$$\mathcal{A}(f) = h_n(f) \times \mathcal{L}(f), \tag{1}$$

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

$$h_n(f) = \frac{1}{n^2}\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ 1 & 1 & \ddots & 1 \\ 1 & 1 & \dots & 1 \end{pmatrix}.$$

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

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

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

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

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

Figure 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 use a simple threshold in the saliency map to detect the

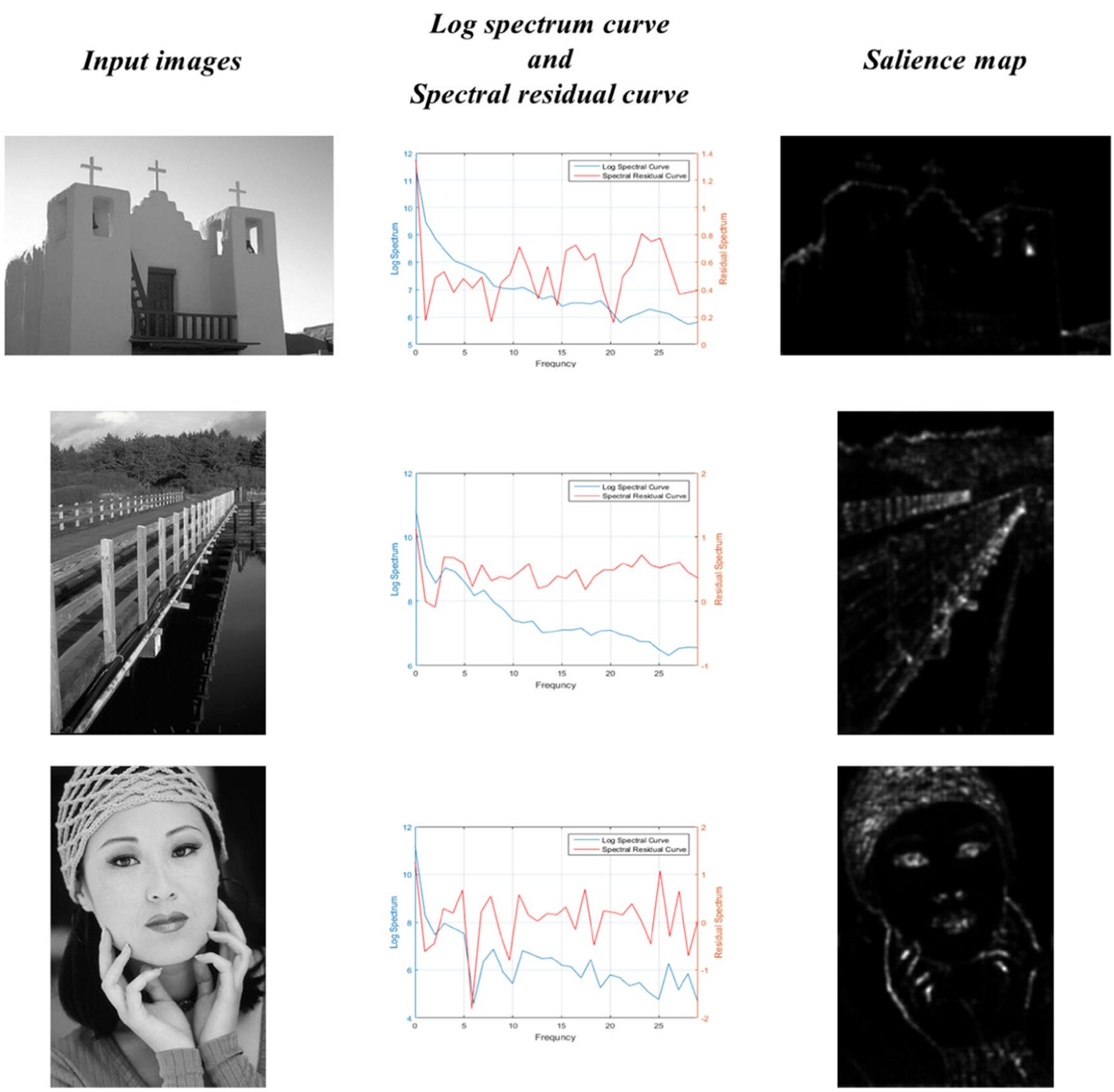

Fig. 1 Saliency map construction. The first column shows the original images. The second column shows the log spectrum (blue solid line) and the spectrum residual (red solid line) of the corresponding image. The saliency maps are on the third column

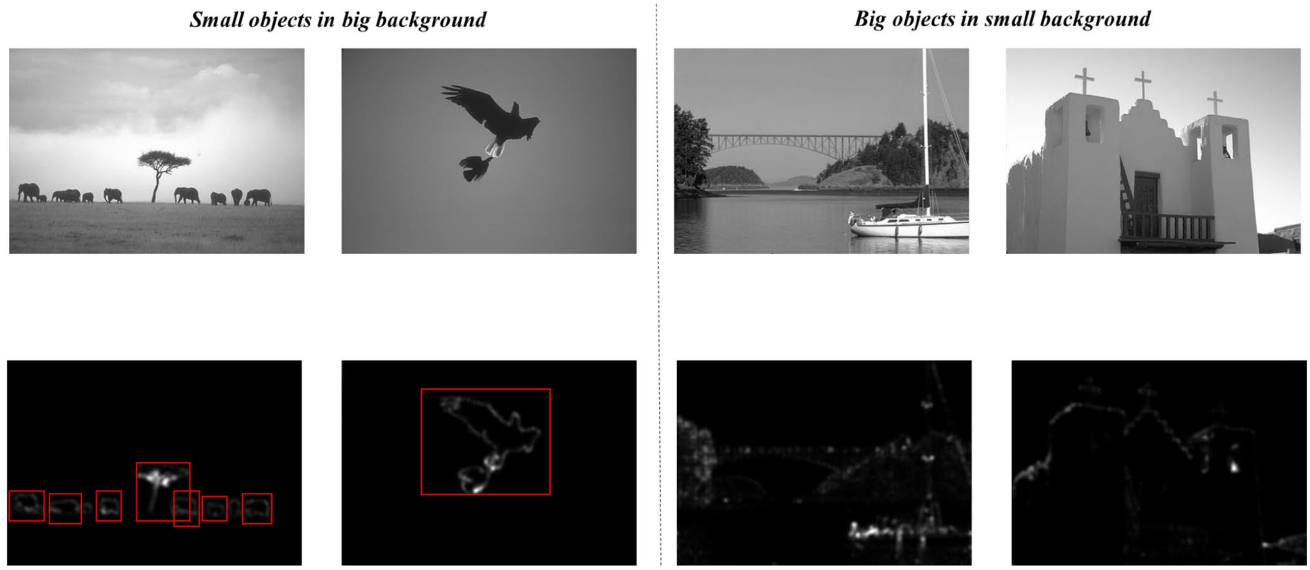

Fig. 2 Proto-objects extraction based on saliency map

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: $i \in \{1, 2, \ldots, N\}$ denotes an image pixel index, x_i is the ith pixel in image, and $j \in \{1, 2, \ldots, L\}$ represents the class label index. In the conventional GMM, the conditional probability that x_i belongs to class j is given by $\Phi(x_i \mid \boldsymbol{\theta}_j)$, which denotes the jth Gaussian distribution component:

$$\boldsymbol{\Phi}(x_i|\boldsymbol{\theta}_{\rm j}) = \frac{1}{2\pi|\Sigma_j|}\exp\left[-\frac{1}{2}(x_i-\mu_j)^T\Sigma_j^{-1}(x_i-\mu_j)\right], \tag{4}$$

where $\boldsymbol{\theta}_j = \{\mu_j, \Sigma_{\rm j}\}$ denotes the mean and the variance of the jth Gaussian distribution.

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

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

where $\boldsymbol{\Pi} = \{\pi_1,\pi_2,\ldots,\pi_L\}$ denotes the set of prior probabilities (also called mixture component weights) and $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_L\}$ is the set of parameters of all Gaussian distributions.

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

$$f(x_i|\boldsymbol{\Psi}) = \sum_{j=1}^{L}\pi_{ij}\left[\sum_{m\in\mathcal{N}_i}\frac{S(x_m)}{R_i}p(x_m|\boldsymbol{\Theta}_j)\right], \tag{6}$$

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 $\sum_{j=1}^{L}\pi_{ij} = 1$; $\mathcal{N}_i$ denotes the neighborhood of the pixel x_i; R_i is the sum of the saliency map values inside $\mathcal{N}_i$; $\boldsymbol{\Psi}$ denotes the parameters set containing all the parameters $\boldsymbol{\Psi} = \{\pi_{11}, \pi_{12},\ldots, \pi_{1L}, \pi_{21}, \pi_{22},\ldots, \pi_{2L}, \pi_{N1}, \pi_{N2},\ldots, \pi_{NL}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2,\ldots, \boldsymbol{\theta}_L\}$; and $S(x_m)$ is the saliency map value at location x_m.

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

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

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

$$\gamma_{ij}^{(t)} = \frac{\pi_{ij}^{(t)} \sum_{m \in \mathcal{N}_i} \frac{\mathcal{S}(x_m)}{R_i} p\left(x_m | \theta_j^{(t)}\right)}{\sum_{h=1}^{L} \pi_{ih}^{(t)} \sum_{m \in \mathcal{N}_i} \frac{\mathcal{S}(x_m)}{R_i} p\left(x_m | \theta_h^{(t)}\right)}. \tag{8}$$

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

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

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

And the prior probability is given by:

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

2.3 Flowchart of GMM-SMSI

The flowchart of the proposed model is described as follows:
Step 1: Salience map extraction.

a. The image is converted to the spectral domain using *FFT*. This gives the amplitude spectrum $\mathcal{F}(f)$ and the phase spectrum $\mathcal{P}(f)$ of the image.
b. The log spectrum representation $\mathcal{L}(f)$ is given by the logarithm of $\mathcal{F}(f)$.
c. The estimation of the average spectrum $\mathcal{A}(f)$ is given by using (1).
d. The calculation of the residual value $\mathcal{R}(f)$ is given by using (2).
e. The generation of the saliency map $\mathcal{S}(x)$ is given by using (3).

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

a. The *k*-means algorithm is used to initialize the parameters sets $\mathbf{\Psi^{(0)}}$.
b. The EM algorithm (Eqs. (8)–(11)) is applied for the parameters estimation until convergence. At the end, we get the parameters set $\mathbf{\Psi(c)}$.
c. The image pixels are then labeled based on a highest *posterior* probability.

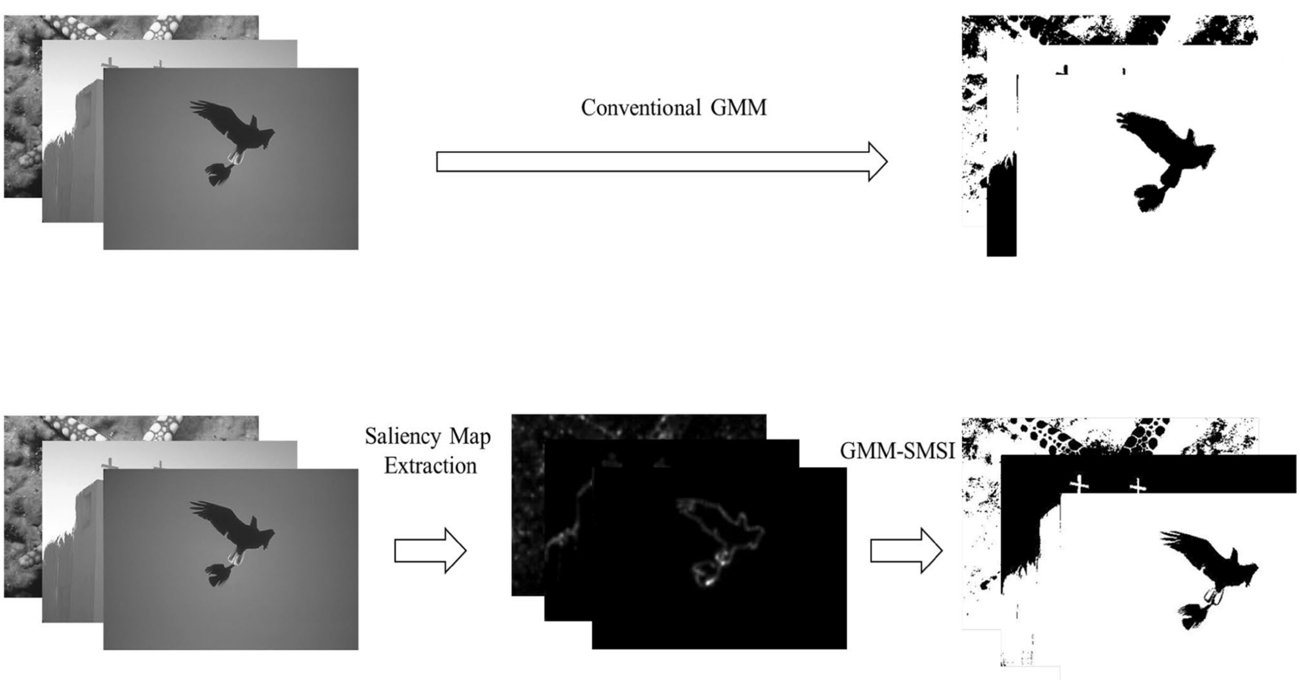

Fig. 3 Flowchart of the conventional GMM and GMM-SMSI

A brief example of the flowchart of the conventional GMM and GMM-SMSI is 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 incorporation of an arithmetic mean template to both the conditional and prior probabilities of a GMM (ACAP) [23]. Our experiments have been performed on MATLAB R2013a and executed on an Intel i5 Core 2.8 GHz CPU with 12.0 GB RAM. We experimentally evaluate 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 that 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 images content: a tiny object in a large background region, a large object in small background region, buildings, and human face images. Then we globally compared the overall performance of the several methods together using the average PR index and the average computation time when the methods were applied on all the images of the *Berkeley* image dataset.

3.1 Tiny object in large background region

In the first experiment, we chose an image (481×321) with two birds in the sky in order to show the ability to segment tiny objects in a large background region (Fig. 4a). 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. 4b, PR = 0.9835), there was a large misclassification of the sky and also of the region between the birds. FLICM (Fig. 4c, PR = 0.9834) showed a similar misclassification of the sky; however, the two birds were now disconnected. In HMRF-FCM (Fig. 4d, PR = 0.9853), the sky misclassification was smaller than in SVFMM and FLICM; however, the two birds were difficult to separate. The accuracy of the segmentation for ACAP (Fig. 4e, PR = 0.9855) was better than that for the other three methods because there was a good classification of sky; however, the two birds were not separated. Our method, GMM-SMSI (Fig. 4f, PR = 0.9864), was able to distinguish the two 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.

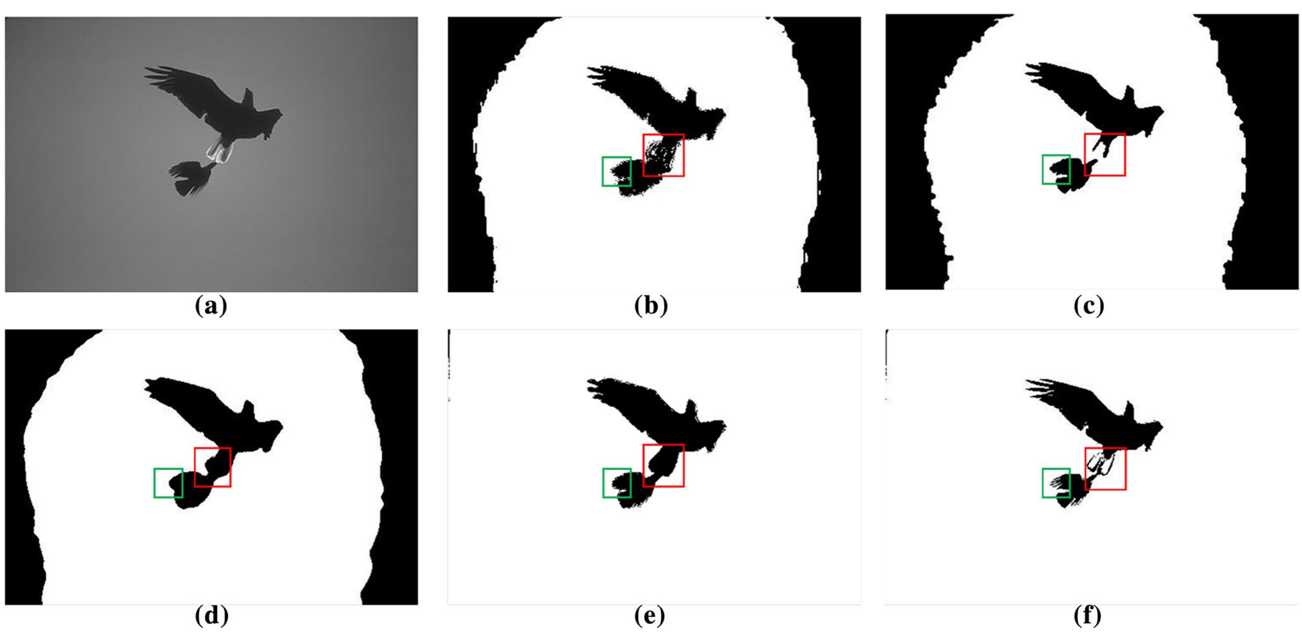

Fig. 4 Tiny object in a large background region segmentation results. **a** Original image. **b** SVFMM, PR = 0.9835. **c** FLICM, PR = 0.9834. **d** HMRF-FCM, PR = 0.9853. **e** ACAP, PR = 0.9855. **f** GMM-SMSI, PR = 0.9864

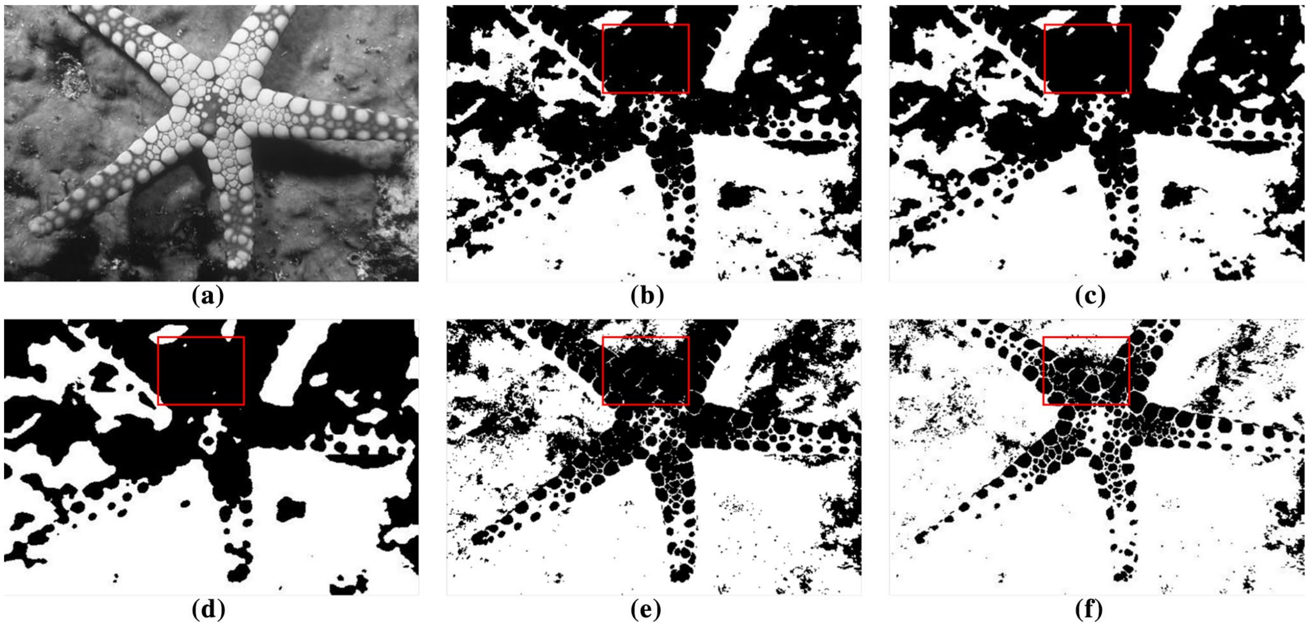

Fig. 5 Large object in a small background region segmentation results. **a** Original image. **b** SVFMM, PR = 0.6835. **c** FLICM, PR = 0.6834. **d** HMRF-FCM, PR = 0.7853. **e** ACAP, PR = 0.6855. **f** GMM-SMSI, PR = 0.6986

3.2 Large object in small background region

In the second experiment, we chose an image (481 × 321) with a relatively large starfish in the seabed (Fig. 5a). We also tried to segment the image into two classes: the background and the starfish. As shown in Fig. 5b, the segmentation for SVFMM (PR = 0.6835) was not able not to distinguish the upper part of the starfish from the background. It is shown in Fig. 5c that FLICM (PR = 0. 6834) had a similar behavior compared with SVFMM. Figure 5d 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 for SVFMM, FLICM and HMRF-FCM, as shown in Fig. 5e, and the starfish was clearly separated from the background. There were still misclassifications of pixels at the edge of the starfish. It can be noticed (Fig. 5f) 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.

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. 6a). It is shown in Fig. 6b–d that for SVFMM (PR = 0. 7204), FLICM (PR = 0. 8977) and HMRF-FCM (PR = 0. 8379), the top right corner of the background was misclassified as church region. The misclassification of SVFMM was much larger than that of FLICM and HMRF-FCM. The segmentation accuracy of ACAP (PR = 0.8599) was better, and the sky and the church were clearly separated, as Fig. 6e shows. The segmentation of our method (PR = 0. 8362) showed more details in the church, such as stairs, the left window and the door (Fig. 6f). It was more in phase with human vision. However, it has to be noted that the PR index value of our method was lower than that of 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. 7a. The goal was to segment the image into two classes. In Fig. 7b, 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. 7c, FLICM (PR = 0. 8977) provided correct facial information; however, the textures of the clothes were not clear. HMRF-FCM (PR = 0. 8379) achieved a better clustering than SVFMM and FLICM but with a lot of lost information, as shown in Fig. 7d. Similar to HMRF-FCM, ACAP (PR = 0. 8599) also achieved better clustering, as shown in Fig. 7e. In Fig. 7f, it can be seen that our algorithm (PR = 0.8362) showed

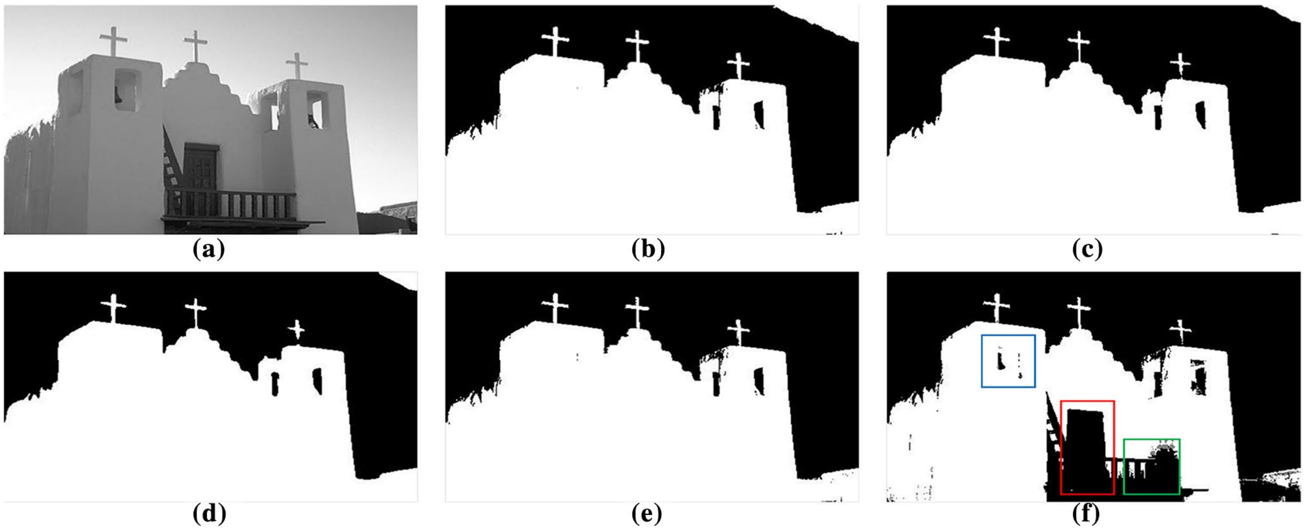

Fig. 6 Building image segmentation results. **a** Original image. **b** SVFMM, PR = 0.7204. **c** FLICM, PR = 0.8977. **d** HMRF-FCM, PR = 0.8379. **e** ACAP, PR = 0.8599. **f** GMM-SMSI, PR = 0.8362

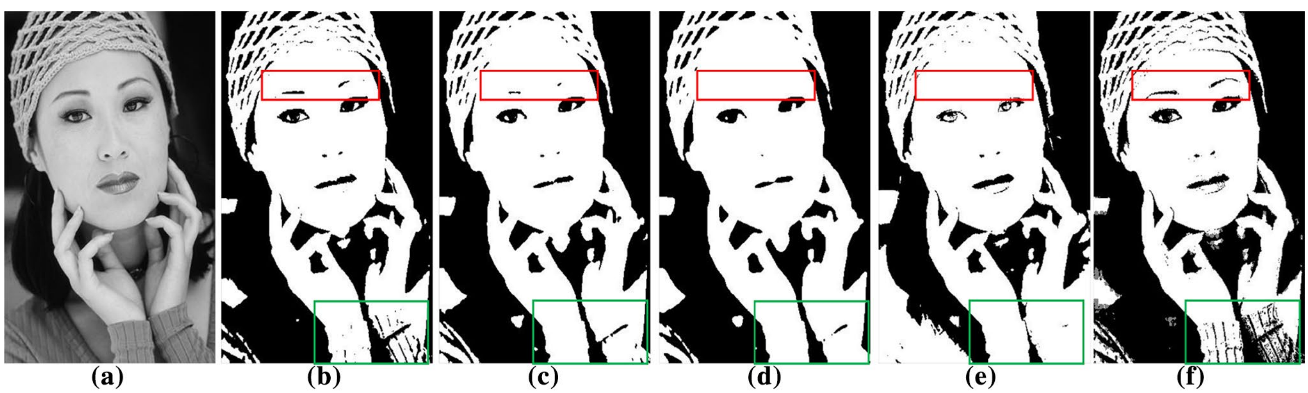

Fig. 7 Face image segmentation results. **a** Original image. **b** SVFMM, PR = 0.7204. **c** FLICM, PR = 0.8977. **d** HMRF-FCM, PR = 0.8379. **e** ACAP, PR = 0.8599. **f** GMM-SMSI, PR = 0.8362

more details of the human face, such as the full eyebrow information. It was also true when considering the texture of the clothes. It was noticed that the PR index value of our method was lower than that of 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 Segmentation performance

In this subsection some objective ways to evaluate SVFMM, FLICM, HMRF-FCM, ACAP and GMM-SMSI are proposed. Table 2 presents the PR index values obtained on a sample of nine 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 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. Figure 8 shows the boxplot of the PR indexes obtained by each method on the whole *Berkeley* image dataset. It confirms that GMM-SMSI has 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, HMRF-FCM, ACAP and GMM-SMSI.

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]	HMRF-FCM [18]	ACAP [23]	GMM-SMSI
78019	7	0.7790	0.6580	0.8308	0.8006	**0.8091**
106,025	4	0.6347	0.8116	0.7988	0.8205	**0.8524**
253,036	4	0.6442	0.8838	0.8690	0.9133	**0.9617**
24,063	4	0.7204	**0.8977**	0.8372	0.8599	**0.8362**
61,086	5	0.7143	0.6873	0.7299	0.7344	**0.8034**
22,090	4	0.7752	0.7675	0.7777	0.8004	**0.8026**
302,003	3	0.7170	0.7172	0.7169	0.7179	**0.7816**
Mean PR value of 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**

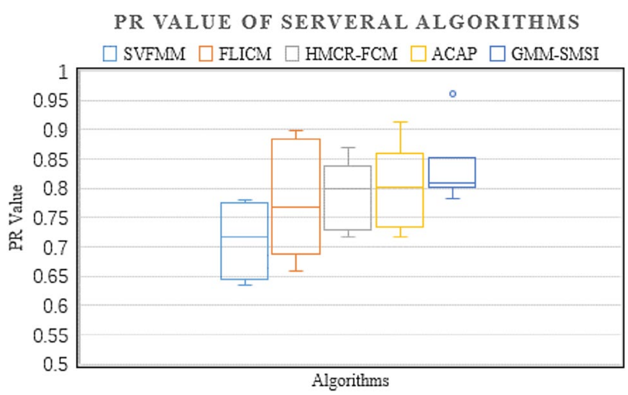

Fig. 8 Boxplot of the PR for several algorithms applied on the *Berkeley* image dataset

Table 2 also presents the average computation time of each method when applied to the whole image set. GMM-SMSI took the lowest computation time that is nearly 14.67% of this of HMRF-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 pixel's neighborhood to enhance the role of significant pixels. Since the saliency map extraction was independent of GMM, it makes the proposed model simple to implement. In addition, the parameters of GMM-SMSI can be easily estimated by expectation–maximization (EM) algorithm. In experiments performed on the public *Berkeley* database, we have demonstrated that the proposed GMM-SMSI method outperformed the state-of-the-art methods in aspects of both classification accuracy and computation time. Moreover, these experiments indicated that our method can detect more object 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.

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

References

1. Sonka M, Hlavac V, Boyle R (2014) Image processing, analysis, and machine vision. Cengage Learning, Boston
2. McLachlan G, Peel D (2004) Finite mixture models. Wiley, New York
3. Bishop CM (2006) Pattern recognition and machine Learning 128
4. Bouguila N (2011) Count data modeling and classification using finite mixtures of distributions. IEEE Trans Neural Netw 22(2):186–198
5. Yuksel SE, Wilson JN, Gader PD (2012) Twenty years of mixture of experts. IEEE Trans Neural Netw Learn Syst 23(8):1177–1193
6. Allili MS, Bouguila N, Ziou D (2007) A robust video foreground segmentation by using generalized gaussian mixture modeling. In: Fourth Canadian conference on computer and robot vision, 2007. CRV'07. IEEE, pp 503–509
7. Allili MS, Bouguila N, Ziou D (2007) Finite generalized Gaussian mixture modeling and applications to image and video foreground

segmentation. In: Fourth Canadian conference on computer and robot vision, 2007. CRV'07. IEEE, pp 183–190
8. Bouwmans T, El Baf F, Vachon B (2008) Background modeling using mixture of gaussians for foreground detection-a survey. Recent Patents Comput Sci 1(3):219–237
9. El Baf F, Bouwmans T, Vachon B (2008) Type-2 fuzzy mixture of Gaussians model: application to background modeling. In: International symposium on visual computing. Springer, pp 772–781
10. Shah M, Deng J, Woodford B (2012) Illumination invariant background model using mixture of Gaussians and SURF features. In: Asian Conference on Computer Vision. Springer, pp 308–314
11. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B (Methodolog) 39:1–38
12. Bilmes JA (1998) A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Int Comput Sci Inst 4(510):126
13. Denœux T (2011) Maximum likelihood estimation from fuzzy data using the EM algorithm. Fuzzy Sets Syst 183(1):72–91
14. McLachlan G, Krishnan T (2007) The EM algorithm and extensions, vol 382. Wiley, New York
15. Figueiredo MAT, Jain AK (2002) Unsupervised learning of finite mixture models. IEEE Trans Pattern Anal Mach Intell 24(3):381–396
16. Blekas K, Likas A, Galatsanos NP, Lagaris IE (2005) A spatially constrained mixture model for image segmentation. IEEE Trans Neural Netw 16(2):494–498
17. Nguyen TM, Wu QJ (2012) Gaussian-mixture-model-based spatial neighborhood relationships for pixel labeling problem. IEEE Trans Syst Man Cybern Part B (Cybernetics) 42(1):193–202
18. Chatzis SP, Varvarigou TA (2008) A fuzzy clustering approach toward hidden Markov random field models for enhanced spatially constrained image segmentation. IEEE Trans Fuzzy Syst 16(5):1351–1361
19. Diplaros A, Vlassis N, Gevers T (2007) A spatially constrained generative model and an EM algorithm for image segmentation. IEEE Trans Neural Netw 18(3):798–808
20. Sanjay-Gopal S, Hebert TJ (1998) Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm. IEEE Trans Image Process 7(7):1014–1028
21. Tang H, Dillenseger J-L, Bao XD, Luo LM (2009) A vectorial image soft segmentation method based on neighborhood weighted Gaussian mixture model. Comput Med Imaging Graph 33(8):644–650
22. Zhang H, Wu QJ, Nguyen TM (2013) Image segmentation by a robust modified gaussian mixture model. In: 2013 IEEE international conference on acoustics, speech and signal processing. IEEE, pp 1478–1482
23. Zhang H, Wu QJ, Nguyen TM (2013) Incorporating mean template into finite mixture model for image segmentation. IEEE Trans Neural Netw Learn Syst 24(2):328–335
24. Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cogn Psychol 12(1):97–136
25. Cheng M-M, Mitra NJ, Huang X, Torr PH, Hu S-M (2015) Global contrast based salient region detection. IEEE Trans Pattern Anal Mach Intell 37(3):569–582
26. Rensink RA, O'Regan JK, Clark JJ (1997) To see or not to see: the need for attention to perceive changes in scenes. Psychol Sci 8(5):368–373
27. Rensink RA, Enns JT (1995) Preemption effects in visual search: evidence for low-level grouping. Psychol Rev 102(1):101
28. Rensink RA (2000) Seeing, sensing, and scrutinizing. Vision Res 40(10):1469–1487
29. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259
30. Borji A, Itti L (2013) State-of-the-art in visual attention modeling. IEEE Trans Pattern Anal Mach Intell 35(1):185–207
31. Walther D, Itti L, Riesenhuber M, Poggio T, Koch C (2002) Attentional selection for object recognition—a gentle way. In: International workshop on biologically motivated computer vision. Springer, pp 472–479
32. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. In: 2007 IEEE conference on computer vision and pattern recognition. IEEE, pp 1–8
33. Ruderman DL (1994) The statistics of natural images. Netw Comput Neural Syst 5(4):517–548
34. Srivastava A, Lee AB, Simoncelli EP, Zhu S-C (2003) On advances in statistical modeling of natural images. J Math Imag Vis 18(1):17–33
35. Krinidis S, Chatzis V (2010) A robust fuzzy local information C-means clustering algorithm. IEEE Trans Image Process 19(5):1328–1337
36. Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of eighth IEEE international conference on computer vision, 2001. ICCV 2001. IEEE, pp 416–423
37. Unnikrishnan R, Pantofaru C, Hebert M (2007) Toward objective evaluation of image segmentation algorithms. IEEE Trans Pattern Anal Mach Intell 29(6):929–944