Neural Comput & Applic (2016) 27:45–58
DOI 10.1007/s00521-013-1544-2

EXTREME LEARNING MACHINE AND APPLICATIONS

Principal pixel analysis and SVM for automatic image segmentation

Xuefei Bai · Wenjian Wang

Received: 31 October 2013 / Accepted: 25 December 2013 / Published online: 14 January 2014

Abstract Segmenting objects from images is an important but highly challenging problem in computer vision and image processing. This paper presents an automatic object segmentation approach based on principal pixel analysis (PPA) and support vector machine (SVM), namely PPA–SVM. The method comprises three main steps: salient region extraction, principal pixel analysis, as well as SVM training and segmentation. We consider global saliency information and color feature by means of visual saliency detection and histogram analysis, such that SVM training data can be selected automatically. Experiment results on a public benchmark dataset demonstrate that, compared with some classical segmentation algorithms, the proposed PPA–SVM method can effectively segment the whole salient object with reasonable better performance and faster speed.

Keywords Image segmentation · Support vector machine · Principal pixel analysis · Visual saliency detection · Training sample selection

1 Introduction

Image segmentation is an important but highly challenging problem in the research field of computer vision and image processing. And this low-level vision task is often used as an indispensable preliminary step in many video and image applications, such as content-based image retrieval [1], image compression [2], robotic application [3] and segmentation in video sequences [4]. In the past few decades, many approaches have been proposed to segment distinct and homogeneous objects from a given image. Among them some methods are closely related with visual attention detection mechanism [5], which is the process of selecting visual information based on saliency in the image itself or on prior knowledge about scenes. Theories of human attention hypothesize that the human vision system only processes parts of an image in detail, while leaving others nearly unprocessed [6]. The selected portion of the image is supposed to represent the most conspicuous parts of the image, known as *conspicuous map* or *saliency map*.

At present, there are many methods for generating saliency map, among which the most influential model was introduced by Itti [7]. Itti's model was inspired biologically and involved three steps: multiple feature extraction, contrast computation and saliency map combination. First, various features like color, intensity, and orientation at different scales were calculated, respectively, and then, a single conspicuous map was formed by applying the center–surround operation across scales to compute the contrast. Finally, conspicuous maps over different scales in different feature space were combined to obtain the saliency map. Afterward many following saliency models adopted the same or similar architecture [8–11]. In addition, other methods for saliency map also include contrast-based algorithms [6, 12, 13], frequency analysis-based algorithms [14–16], hybrid algorithm using edge detection, threshold and distance transform [17], new algorithms based on learning models [18, 19], etc.

X. Bai · W. Wang
School of Computer and Information Technology,
Shanxi University, Taiyuan, People's Republic of China
e-mail: baixuefei@sxu.edu.cn

W. Wang (✉)
Key Laboratory of Computational Intelligence and Chinese
Information Processing of Ministry of Education,
Shanxi University, Taiyuan, People's Republic of China
e-mail: wjwang@sxu.edu.cn

At the same time, various saliency maps have been previously employed for unsupervised object detection and segmentation. Ouerhani et al. [20] exploited the usefulness of visual attention in the seed selection process for segmenting outdoor road traffic scenes. They computed saliency maps through purely data-driven visual attention processing, and then, attention points were provided to be seeds in a seeded region growing algorithm using a color homogeneity criterion. Han et al. [21] proposed an unsupervised extraction model of viewer's attentive objects in color images, in which a classical computational visual attention mechanism and some region growing techniques were integrated to formulate the attention objects as a Markov Random Field. Their model does not work well when there are many attention objects or complicated background exists in the image. Unlike previous segmentation methods only based on saliency maps, Ko et al. [22] developed a novel algorithm for segmenting an object-of-interest (OOI) using both a saliency map and saliency points, which constitute an attention window. And within the attention window, region clustering algorithm and support vector machine (SVM) were used to determine whether a region was part of the OOI or the background. Later, Achanta et al. [23] detected the salient regions by using a contrast determination filter that operates at various scales to generate multiple saliency maps. Then, these individual maps were combined to result in final saliency map, on which some simple segmentation technique can be used to detect the salient region. An unsupervised color segmentation method based on saliency was proposed in [24], whose main idea was to segment the input image several times, and in each time a different salient part of the image was processed. And then, all obtained results were merged into one composite segmentation. The saliency map used in their method was calculated according to local color and texture models. Liu et al. [25] firstly constructed region/boundary saliency maps based on a pre-segmentation result by watershed transform, and then entropy thresholding and flood filling were exploited to generate region/boundary masks. Next, a trimap containing seed regions for attended objects, unattended regions and uncertain regions was obtained as the input of an image matting model, which was utilized to classify the pixels in the uncertain regions to get an accurate attended object segmentation result based on the estimated alpha matte. Lately, a saliency-directed color image segmentation approach using particle swarm optimization was proposed in [26], in which the visual attention saliency map was generated by three (color, intensity and orientation) feature maps, and a hybrid fitness function was adopted for image segmentation.

Besides of the above methods based on saliency map, another kind of methods for object segmentation are using some classical image segmentation algorithms according to the characteristics of images. Liu et al. [27] proposed an unsupervised salient object extraction and segmentation approach using region merging and binary partition tree. Starting from an over-segmentation color image, region merging was performed to generate a binary partition tree, from which an appropriate subset of modes was selected to obtain the contours of salient objects [27]. Different from previous approaches, in order to extract the center of interest from pictures, Zhu et al. [28] developed a two-level segmentation strategy based on photographic theory, which was similar to the mode of how people perceive pictures. Recently, Fu et al. [29] proposed a novel salient object detection and segmentation method based on superpixel.

Although these methods mentioned above can segment salient object from images, there are still some shortcomings. First, some visual attention models and saliency maps used for segmentation need a few tunable parameters and the computational complexity maybe high for practical applications [20–22]. Second, some methods only obtain an attention view or some shape information such as contours or boundaries of the salient region but not segmenting the homogeneous salient object [27, 28]. And these results may yield limited help for further object recognition and image understanding. Therefore, further studies are still necessary.

In recent years, support vector machine [30] has already been applied in image segmentation due to its excellent learning and generalization performance in solving binary classification problems. Yu et.al [31] introduced a new SVM-based approach named fast support vector machine (FSVM) for image segmentation. Pixels in a small part of an object and background were respectively specified by users as training samples of FSVM. Then, a pruning strategy was used to eliminate redundant training vectors. Finally, the remaining pixels were viewed as test data and were segmented into several regions by the trained FSVM classifier. Because training samples of FSVM are prespecified by users, which leads to manual intervention. Furthermore, different training samples will affect the final classification performance of FSVM. A new approach for color image segmentation using SVM and FCM was proposed in [32], in which training samples of SVM were randomly selected by the FCM clustering algorithm. However, the number of clusters for FCM must be set in advance, and the random selection of training samples will also affect the final segmentation performance. Furthermore, although much effort has been put on SVM-based image segmentation and many other approaches have been proposed [33, 34], it is still a challenging task to automatically segment natural images due to their inherent complexity. A crucial problem is about SVM training samples selection, which often depends on human intervention completely or partially, such as methods introduced previously [31, 32]. However unfortunately, manual intervention is not so feasible in most practical applications. But for salient object segmentation, the utilization of visual saliency could provide some useful cues for SVM training samples automatic selection. Therefore, how to effectively exploit visual saliency of image itself to automatically select SVM training samples, as well as explore the excellent classification performance of SVM for salient object segmentation, is the main focus of this paper.

In order to solve the above problems, an automatic approach is needed, which deals with two tasks about salient object: auto-location by visual saliency detection and auto-segmentation based on SVM. We observe that in salient object and background of most images, there always exist some pixels that can represent the majority features of the homogeneous regions where they locate, referred to as "Principal Pixel" herein. Therefore, this paper aims to utilize the principal pixels derived from visual saliency detection and SVM classifier, to automatically locate and segment salient object from images. Compared with other existing methods, the proposed PPA–SVM method offers the following advantages: it is independent of image features such as intensity, shape, texture or other prior knowledge of the given image; the segmentation processing is automatic and adaptive, avoiding any human intervention; training sets and training pixels of SVM are selected in accordance with the characteristics of the image.

The rest of this paper is organized as follows. In Sect. 2, the proposed PPA–SVM method is described in detail, including visual saliency detection, SVM training data generation, training samples selection, along with SVM training and salient object segmentation. Experimental results and analysis are discussed in Sect. 3. Finally, conclusions are addressed in the last section.

2 The proposed PPA–SVM approach

We define principal pixels of an image in terms of spatial location and color feature, referring to some representative pixels that in salient object R_o or background R_b, and with dominant colors of these two regions. So, the proposed PPA–SVM method starts with visual saliency detection to find the prominent locations of the salient object, which results in a coarse partition of salient region R_o and background R_b. And the dominant colors of R_o and R_b, which can represent the distinguishing characteristic of each region, are determined through an adaptive histogram peak selection method. Afterward, SVM-positive and SVM-negative training sets, consisting of principal pixels, are generated respectively. And then, a local homogeneity criterion is proposed for choosing some principal pixels from two training sets as training samples, to indicate the distribution of saliency, spatial and color features of image training data. Finally, a SVM model is trained to segment the salient object from the given image.

The overall procedure of the proposed PPA–SVM method is illustrated in Fig. 1. And the main steps will be explained in the following subsections.

2.1 Salient region detection and extraction

In this subsection, we focus on the automatic detection and extraction of salient region where the salient object maybe locate. Similar to some approaches mentioned previously, visual attention mechanism is adopted in the proposed PPA–SVM method. Human visual attention is one of the most intelligent ability to rapidly detect interesting parts of a given image [5].

Castleman [35] pointed out that the amplitude spectrum specifies how much of each sinusoidal component is present and the phase information specifies where each of the sinusoidal component appears in the image. The location with less periodicity or less homogeneity in vertical or horizonal orientation creates the "pop-out" *proto objects* in the reconstruction of the image, which indicates where the salient object candidates are located, and the phase spectrum of image Fourier transform is the key to calculate the location of salient region [2, 15]. In this paper, similar idea is adopted as the saliency detection scheme due to its low computational cost, full resolution and unsupervised manner. For a given image $I(x, y)$, the saliency map $SM(x, y)$ is calculated as follows:

$$f(x,y) = F(I(x,y)) \tag{1}$$

$$p(x,y) = P(f(x,y)) \tag{2}$$

$$SM(x,y) = g * \|F^{-1}[e^{i \cdot p(x,y)}]\|^2 \tag{3}$$

where F and F^{-1} refer to the Fourier Transform and Inverse Fourier Transform, respectively. $p(x, y)$ represents the phase spectrum of the image, and g is a 2D Gaussian filter ($\sigma = 8$) for a better visual effect as used in Refs. [2, 14, 15].

Saliency map generated in this way like [2, 15] is a gray-level image, which represents the salient value of each pixel. The salient values range from 0 to 255, and the larger the salient value, the more likely the pixel attract observers' interest, as shown in Fig. 2a and b. In order to further identify the location of salient object, the object mask $OM(x, y)$ can be obtained by a binarization precessing of $SM(x, y)$:

$$OM(x,y) = \begin{cases} 0 & \text{if} \quad SM(x,y) < t \\ 1 & \text{if} \quad SM(x,y) \geq t \end{cases} \tag{4}$$

The binarization threshold t is set to be the value which maximizes the discrimination criterion (σ_B^2/σ_W^2) of two classes (salient object and background), where σ_B^2 is the

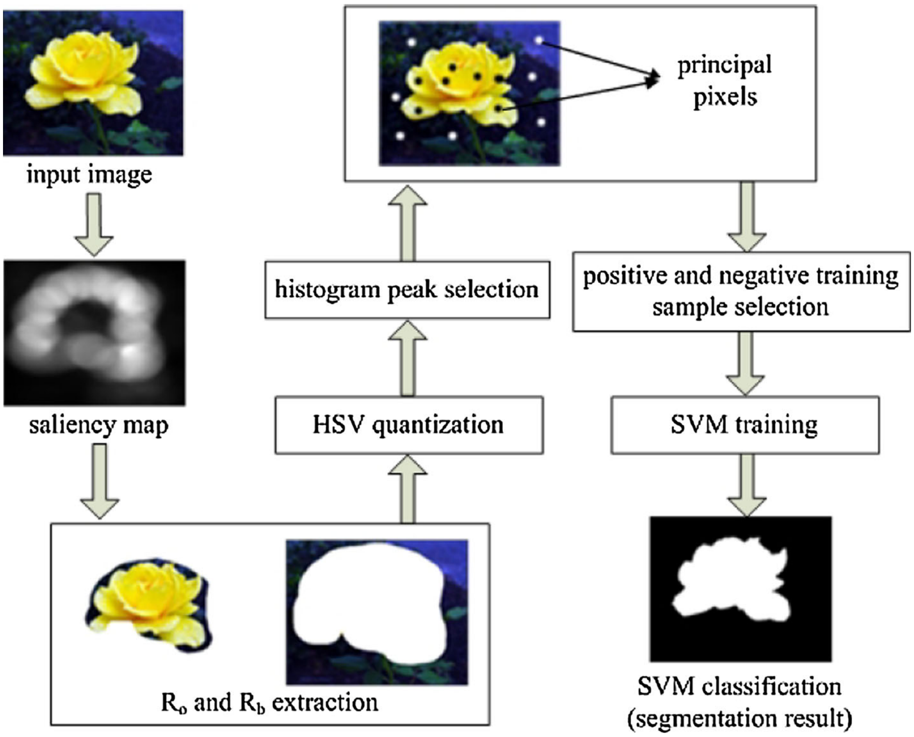
input image
saliency map
R_o and R_b extraction
HSV quantization
histogram peak selection
principal pixels
positive and negative training sample selection
SVM training
SVM classification (segmentation result)

Fig. 1 An overall procedure of the proposed PPA–SVM method

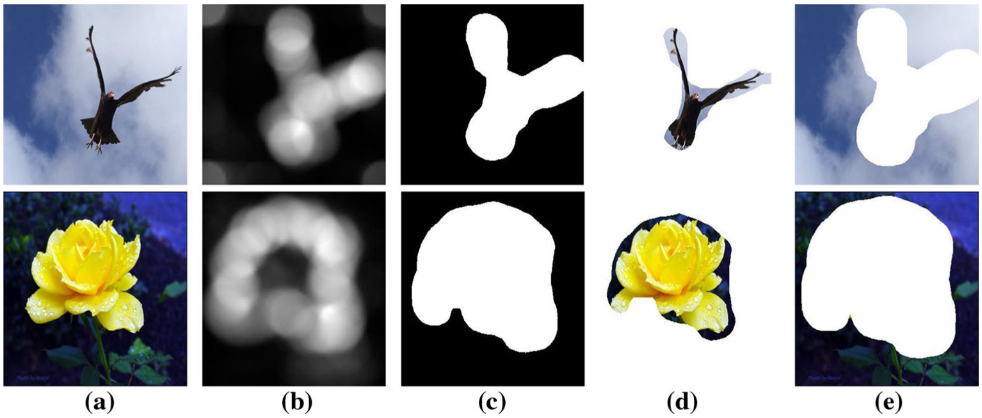

Fig. 2 Some examples of the salient region detection and extraction procedure. **a** Original images, **b** saliency maps, **c** binary masks, **d** salient regions, **e** background

between-class variance and σ_W^2 is the within-class variance, respectively. The binary object mask $OM(x, y)$ obtained is shown in Fig. 2c, where the white area of $OM(x, y)$ represents the rough estimation of salient object S_r, while black area means the rough estimation background B_r. And it can be found that in general natural scene images, most of the salient object appear at or near the center of the image in order to attract attention distinctly, known as center-bias [36]. So, the binary object mask $OM(x, y)$ should be regularized using some morphological operators [37] to remove these uncertain pixels near the boundary of S_r. Here, the boundary of S_r is shrunk to form a more accurate salient object mask M_o, and then it is expanded to form the background mask M_b:

$$\begin{aligned} M_o &= S_r \ominus E_{r_e} \\ M_b &= ((S_r \oplus D_{r_d}) - S_r) \cup B_r \end{aligned} \tag{5}$$

where $\ominus E_{r_e}$ is an erosion operator indicating shrinking region S_r for r_e pixels, and $\oplus D_{r_d}$ is a dilation operator denoting expanding region S_r for r_d pixels. A square structural element with the width of 10 pixels is used in erosion and dilation operators.

Therefore, salient region R_o and background R_b in a given image can be extracted with the masks M_o and M_b, as shown in Fig. 2d and e. Although salient region and background obtained in this stage are only approximate representations, they uniformly highlight most salient object with full resolution. What's more, the extraction procedure is relatively fast, which can help to refine salient object in the subsequent steps.

2.2 Principal pixel set construction

After the coarser partition of salient region and background, we use color feature of R_o and R_b to get more precise segmentation as color is a commonly used feature. However, the most popular RGB color space contains 256^3 possible colors, which is too computationally expensive in feature extraction process. On the other hand, HSV (Hue: [0, 360°], Saturation: [0, 1], and Value: [0, 1]), which is capable of emphasizing human visual perception, is shown to have better results for image segmentation than RGB color space [38]. Thus, in order to extract the color feature of R_o and R_b, a quantization operation in HSV color space is firstly implemented to reduce the computational complexity. Accordingly for regions R_o and R_b, each channel of HSV color space is quantized to different values, and then, a one-dimensional histogram is generated, respectively. Next, the most frequently occurring colors in R_o and R_b, i.e., dominant colors, which can distinguish salient object and background, are chosen by histogram peak selection.

Because the human visual system is more sensitive to hue than to saturation and intensity so that the hue channel should be quantized finer than saturation and intensity [38]. And it is well known that the color distribution (red, orange, yellow, green, cyan, blue and purple) of the hue channel is not uniform; therefore, similar to [39], a non-uniform quantization scheme is applied here. As a result, the hue channel is quantized to 7 non-uniform bins represented from 0 to 6, and each indicates a major color. And the saturation and intensity channels are quantized non-uniformly to 3 bins in the same way. The quantization scheme can also be summarized as follows:

$$H = \begin{cases} 0 & \text{if } h \in (342, 16] \\ 1 & \text{if } h \in (16, 42] \\ 2 & \text{if } h \in (42, 64] \\ 3 & \text{if } h \in (64, 152] \\ 4 & \text{if } h \in (152, 195] \\ 5 & \text{if } h \in (195, 280] \\ 6 & \text{if } h \in (280, 342] \end{cases}$$
$$S = \begin{cases} 0 & \text{if } s \in [0, 0.3) \\ 1 & \text{if } s \in [0.3, 0.8) \\ 2 & \text{if } s \in [0.8, 1] \end{cases} \qquad (6)$$
$$V = \begin{cases} 0 & \text{if } v \in [0, 0.2) \\ 1 & \text{if } v \in [0.2, 0.7) \\ 2 & \text{if } v \in [0.7, 1] \end{cases}$$

According to the above quantization scheme, one-dimensional feature vector is constructed by three channels values as follows

$$L = Q_s Q_v H + Q_v S + V \qquad (7)$$

where Q_s and Q_v are quantization coefficients of saturation and intensity channel, respectively, and they are set as $Q_s = 3$, $Q_v = 3$ in this work; hence,

$$L = 9H + 3S + V \qquad (8)$$

Thus, three channels hue, saturation and intensity can be distributed in one-dimensional vector L and $L \in \{0, 1, \cdots, 62\}$. Because the quantization result has only 63 bins, the computational complexity will be decreased tremendously. Furthermore, by considering the non-uniform character in three channels, the quantization result is more similar to the human vision [39]. So color value of each pixel in R_o and R_b regions can be quantized to one of the 63 colors, and the HSV histograms with 63 bins of R_o and R_b will be calculated to determine their dominant colors by peak selection, respectively.

Generally, the total numbers of dominant color in salient region R_o and background R_b are both limited, our previous statistical results on 1,000 natural images also confirms this point, i.e., no more than three or four dominant colors exist in salient object or background in above 85 % images. Hence, it is assumed in our experiments that no more than three dominant colors are necessary to describe R_o and R_b regions. Taking salient regions R_o for the example, the main steps to adaptively select histogram peaks are briefly stated as follows:

Step 1 Calculate histogram of region R_o after HSV color space quantization:

$$H^o(l_i) = \frac{Num(f(x,y) = l_i)}{Num(R_o)}, (x,y) \in R_o, l_i \in \{0, 1, \ldots, 62\};$$

where $Num(R_o)$ means the total number of pixels in region R_o, and $Num(f(x, y) = l_i)$ is the number of pixels with color level l_i in R_o.

Step 2 Identify all peaks $P_{ko} : P_{l_1}, P_{l_2}, \ldots, P_{l_k}$, l_i is the color level index of ith peak, and $l_1 < l_2 < \cdots < l_k$.

Step 3 Compute the max and min peak values of H^o. $P_{max} = max\{P_{l_1}, P_{l_2}, \ldots, P_{l_k}\}$, $P_{min} = min\{P_{l_1}, P_{l_2}, \ldots, P_{l_k}\}$, the mean value $\mu_m = (P_{max} + P_{min})/2$ and the standard deviation $\sigma_m = \sqrt{\sum_{i=1}^{k} (P_{l_i} - \mu_m)^2 / k}$. The height threshold in R_o is set as $T_{ho} = \mu_m - \sigma_m$. Some lower peaks are removed based on T_{ho}, and new peaks $P_{ho} : P_{l_1}, P_{l_2}, \ldots, P_{l_h}$ are generated.

Step 4 Remove some peaks based on width threshold T_{wo}. The threshold $T_{wo} = 20$ is set based on the assumption that there should be no more than 3 dominant colors in a salient object. For two adjacent peaks P_{l_i} and P_{l_j}, if $(l_j - l_i) < T_{ws}$, then keep the peak with greater value and remove another peak from P_{ho}.

Step 5 Output the final peak sequence P_{no}, and dominant colors of R_o are determined as $C_o : l_1, l_2, \ldots, l_n$.

Similar to the manner of salient region R_o, the dominant colors C_b of background R_b can be obtained too. After the above processing, principal pixels that locate in salient region R_o and with dominant color C_o, as well as locate in background R_b and with dominant color C_b, can be selected. Meanwhile, the positive training set TS_p and negative training set TS_n of SVM can be constructed as:

$$\begin{aligned} TS_p &= \{(x,y)|f(x,y) = i, (x,y) \in R_o, i \in C_o\} \\ TS_n &= \{(x,y)|f(x,y) = i, (x,y) \in R_b, i \in C_b\} \end{aligned} \tag{9}$$

Considering global saliency information, spatial location and local color feature, training data derived from principal pixel analysis have some advantages over existed methods [31, 32]: without human intervention, fully representing image characteristic distribution, strong robustness and computational efficiency.

2.3 SVM training samples selection

Although principal pixels selected from the previous stage can be used to train a SVM classifier, the total numbers of pixels in TS_p or TS_n are too large to be used as SVM training data directly. Furthermore, when mapped to a higher feature space, training pixels with the same color value in a small local area may be redundant to learn the separating hyperplane, so selecting central pixel to replace the surrounding area is a way to reduce redundancy and improve the learning efficiency. Therefore, a neighborhood homogeneity criterion is adopted to select small part of pixels in TS_p and TS_n as training samples of SVM.

For a pixel $p(i, j)$ in training set TS_p or TS_n, its local homogeneity in 3 × 3 neighborhood is measured as:

$$M_p = D_p^{3\times3} = \sum_{q \in N_p^{3\times3}} d(p,q) \tag{10}$$

where $d(p, q)$ is the Euclidean color distance between pixel p and q in quantized HSV color space, $N_p^{3\times3}$ is the pixel set of adjacent eight neighbors of pixel p.

Because the color difference at lower level can indicate more intuitive local homogeneity, principal pixels that meet $M_p \le T_h$ (T_h is the homogeneity threshold) in TS_p and TS_n will be selected as training samples of SVM. The smaller the threshold value, the more pixels ultimately selected to train SVM classifier, and vice versa.

2.4 Feature vector representation

Image segmentation can be regraded as a binary classification task, whose goal is to assign a label to each pixel in order to identify whether it belongs to salient object or background. Here, color, texture, spatial and global saliency information of each pixel are employed as feature vector in SVM training and segmentation.

The pixel-level color feature CF_{xy} of an image pixel p at location (x, y) is represented as:

$$CF_{xy} = (RG_{xy}, BY_{xy}, I_{xy}) \tag{11}$$

where $RG_{xy} = R - G$, $BY_{xy} = B - Y$, and $R = r - (g + b)/2$, $G = g - (r + b)/2$, $B = b - (r + g)/2$, $Y = (r + g)/2 - (r - g)/2 - b$ are four broadly-tuned color tunnels. r, g, b are the red, green and blue components of the pixel p. The intensity feature $I_{xy} = 0.299{*}r + 0.587{*}g + 0.114{*}b$.

To obtain the pixel-level texture feature, Gabor filter is adopted here as in [31]. The pixel-level texture feature TF_{xy} of the image pixel p at location (x, y) is:

$$TF_{xy} = (E_{xy}, G_{xy}) \tag{12}$$

where E_{xy} denotes the maximum of the six coefficients at (x, y) and G_{xy} denotes the maximum of 6 gradient magnitudes at location (x, y).

Another important characteristics of an image is that neighboring pixels are highly correlated. In other words, neighboring pixels always possess similar feature values, and the probability that they belong to the same class is high. Therefore, to exploit the spatial information, two-dimensional coordinates of the pixel are used.

Besides three kinds of features described above, global saliency information Sxy of pixel p, reflecting how strong the pixel can draw the viewer's attention without any prior knowledge, is also considered as feature vector.

In summary, the feature vector of image pixel is expressed as:

$$F_{xy} = (RG_{xy}, BY_{xy}, I_{xy}, E_{xy}, G_{xy}, x, y, Sxy) \tag{13}$$

2.5 SVM for salient object segmentation

For a given image, the proposed PPA–SVM method is summarized as follows:

Step 1 Detect and extract salient region R_o and background R_b as described in Sect. 2.1.

Step 2 Generate training sets TS_p and TS_n of two regions R_o and R_b, respectively, by principal pixel analysis as detailed in Sect. 2.2.

Step 3 Select training samples from training sets TS_p and TS_n according to the neighborhood homogeneity threshold, and extract feature vectors to train SVM model as described in Sects. 2.3 and 2.4

Step 4 Segment out the whole salient object from the given image by trained SVM model.

3 Experiments and discussions

In order to validate the effectiveness of the proposed PPA–SVM method, it is compared with two automatic segmentation methods given by [40, 41] and a saliency map-based segmentation method given by [6].

Kmeans clustering [40] and Ncuts method [41] are two classical approaches adopted to partition the image into some regions. These two methods are unsupervised but to classify pixels into clusters automatically according to the features of the pixels. We compared the proposed PPA–SVM method with these two methods in order to test the automatic segmentation performance. On the other hand, as mentioned before, saliency map is often used to segment salient object from background, and there are a large number of segmentation algorithms based on saliency map. But as far as we know, RCC method-based segmentation [6], which iteratively applies Grabcut [42] to refine the segmentation result initially obtained by thresholding the saliency map, can yield higher precision and better recall and outperforms other existing saliency map-based methods. Hence, we compared the proposed PPA–SVM method with RCC method for the purpose of testing the salient object detection and segmentation performance.

For three compared methods Kmeans, Ncut and RCC, we execute their corresponding public available softwares or codes, in which Kmeans and Ncut are implemented in MATLAB, while RCC in C++. The number of clusters are set to 2 for Kmeans and Ncut methods in our experiments. We use LibSVM toolbox [43] and kernel function is polynomial with $d = 1$ for PPA–SVM training and segmenting in all experiments that implemented in MATLAB. The local homogeneity criterion for selecting training pixels in PPA–SVM is set as $T_h = 0$.

3.1 Experiment dataset and evaluation metrics

We evaluate our proposed PPA–SVM method on a public available image dataset with ground truth segmentation results [16]. The dataset is derived from the MSRA dataset proposed by Liu et al. [13], containing 1,000 images collected mostly from image forums and image search engines. Each image contains at least one salient object or one distinctive foreground object in simple or complex scenes. And these salient objects differ in category, color, shape, size, and so on. In other words, there is no more prior knowledge or constraint on these objects. And accurate object-contour-based ground truth segmentations contained the most salient region for each image are provided too in this dataset. Hence, many saliency models for detecting or segmenting salient object evaluate their performance on this dataset [6, 16]. And the resolution of most images in this database is 400 × 300.

In our experiments, test images are divided into three categories according to the difference degree between salient object and its surrounding regions and the content of salient object:

test 1 Images with higher saliency, i.e., there are distinct feature differences between the salient object and its surrounding regions; meanwhile, color and texture features in salient object and background are largely homogeneous.

test 2 Images with mid-level saliency, i.e., the differences between the salient object and background are not so obvious, or the content of images are relatively complex.

test 3 Images with lower saliency, i.e., salient objects with discontinuous and irregular edges or complex content.

Five evaluation metrics are adopted to quantitatively assess the segmentation performance in the experiment.

1. The segmentation error rate (ER) is defined as:

$$ER = \frac{(N_f + N_m)}{N_t} \times 100\% \tag{14}$$

   where N_f is the number of false-segmented image pixels, N_m denotes the number of miss-segmented image pixels, and N_t is the total number of image pixels.

2. Probabilistic Rand Index (PRI) [44] is commonly used to measure the similarity between two clusterings. In our experiments, it is employed to count the fraction of pairs of pixels whose labels are consistent between the compared segmentation S_c and the ground truth segmentation S_g. PRI is defined as:

$$PRI(S_c, S_g) = \frac{1}{\binom{N}{2}} \sum_{i,j} [c_{ij}p_{ij} + (1 - c_{ij})(1 - p_{ij})] \tag{15}$$

   where N is the number of pixels, and p_{ij} is the ground truth probability that $\prod(l_i = l_j)$, and $c_{ij} = \prod(l_i^{S_c} = l_j^{S_c})$. The PRI has a value in the interval [0, 1], with 0 indicates that the two segmentations do not agree on any pair of pixels and 1 indicates that the compared segmentation S_c is exactly the same as the ground truth segmentation S_g.

3. Variation of information (VI) introduced in [45] measures the distance between two clusterings in terms of the information difference between them. As image segmentation can be seen as a clustering problem, the VI metric is defined as the distance between two segmentations as the average conditional entropy of one segmentation given the another. It can roughly measures the amount of randomness in one segmentation which cannot be explained by the other.

$$VI(S_c, S_g) = H(S_c) + H(S_g) - 2I(S_c, S_g) \quad (16)$$

where H and I represent the entropies and the mutual information between the compared segmentation S_c and ground truth S_g, respectively.

4. Global consistency error (GCE) [46] measures the extent to which one segmentation can be viewed as a refinement of the other. Segmentations which are related in this manner are considered to be consistent, since they could represent the same natural image segmented at different scales. This measure allows for refinement, but suffers from degeneracy. Let $R(S, p_i)$ be the set of pixels in segmentations S that contains pixel p_i, the local refinement error is defined as:

$$E(S_1, S_2, p_i) = \frac{|R(S_1, p_i) \backslash R(S_2, p_i)|}{|R(S_1, p_i)|} \quad (17)$$

This error is not symmetric with respect to the compared segmentations and takes the value of zero when S_1 is a refinement of S_2 at pixel p_i, GCE is then defined as:

$$GCE(S_1, S_2) = \frac{1}{n} min\left\{ \sum_i E(S_1, S_2, p_i), \sum_i E(S_2, S_1, p_i) \right\} \quad (18)$$

5. The CPU time to segment an image.

3.2 Experiment results and analysis

Some visual comparisons of salient object segmentation using Kmeans, Ncut, RCC and PPA–SVM method for higher-saliency test images are shown in Fig. 3. In this experiment, most salient objects in test images can be segmented effectively from background for these four methods. However, the segmentation results of PPA–SVM are the closest to that of ground truth segmentations. Obviously, in the first flower image, both the black center and the stem of the flower are segmented accurately by PPA–SVM, while they are not segmented fully using other three methods. For the second and the third images, Kmeans method can segment most salient objects but with some noises, while segmentation results of Ncut are not accurate enough. Segmentation result of RCC method for the third image is close to the ground truth segmentation, but it lost some edge information.

Figure 4 shows visual comparison results for some mid-level saliency images. Similar to the results in Fig. 3, segmentation results of Ncut method are not accurate enough and usually accompanied with information loss in salient object. Kmeans method can segment most salient objects but with some noises, while segmentation results of PPA–SVM, very closer to the ground truth segmentations, can include more detailed texture and color information of salient objects. Only for the first image, segmentation result of RCC method is the most closest to the ground truth segmentation, but segmentation result of PPA–SVM can provide more help for later image recognition and understanding. For other two images, those segmentation results of RCC are inferior to that of PPA–SVM.

And Fig. 5 shows visual comparison results for some images with lower saliency. Segmentation results of Ncut method contains only some parts of salient object, and Kmeans method produces lower quality segmentation results with noises and discontinuous edges of salient objects. RCC method can segment most salient object in test images. The proposed PPA–SVM method can obtain

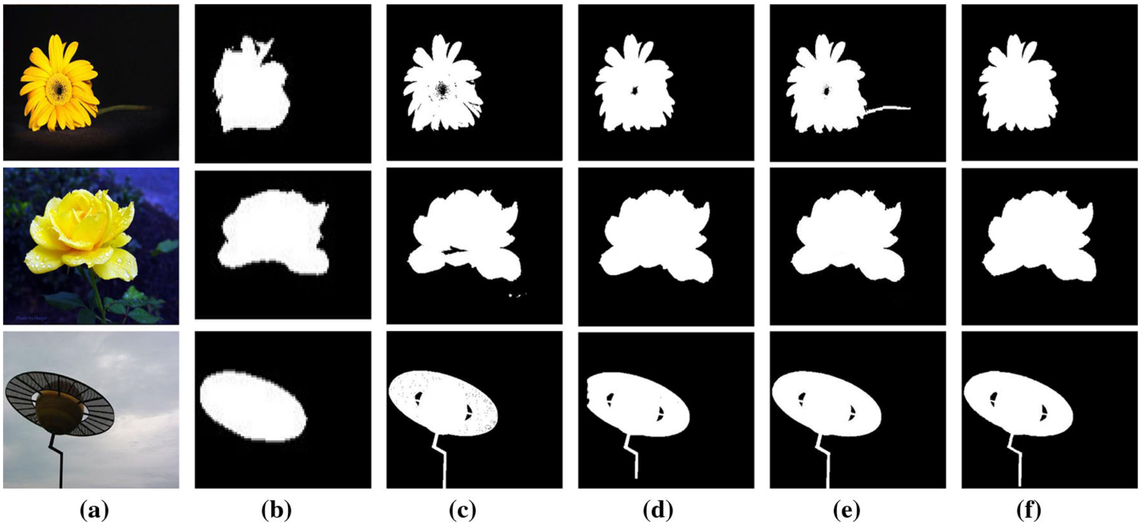

Fig. 3 Comparison of salient object segmentation. **a** Original images; **b** segmentation result using Ncut; **c** segmentation results using Kmeans; **d** segmentation results using RCC; **e** segmentation results of PPA–SVM; **f** ground truth segmentations

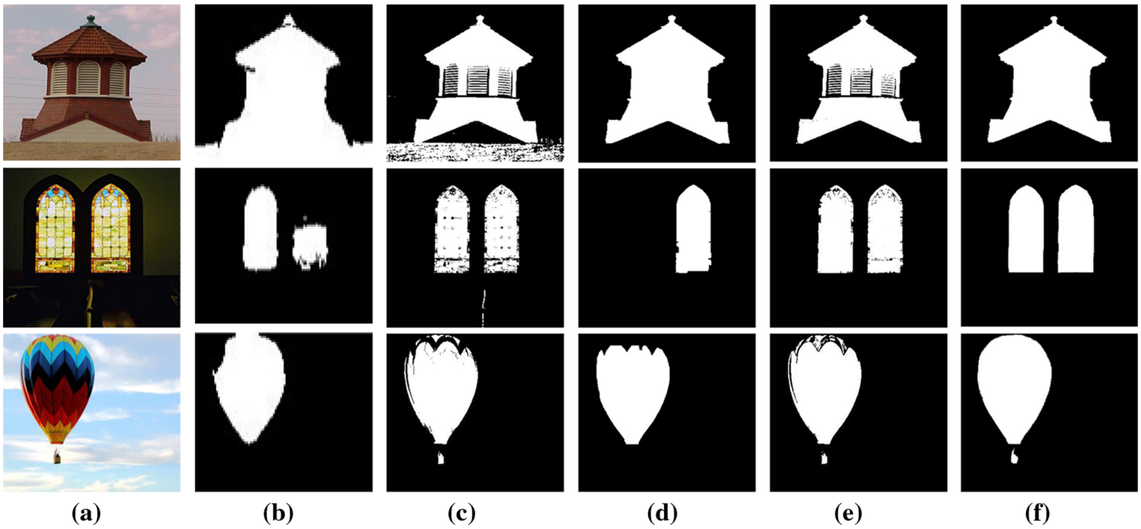

Fig. 4 Comparison of salient object segmentation. **a** Original images; **b** segmentation result using Ncut; **c** segmentation results using Kmeans; **d** segmentation results using RCC; **e** segmentation results of PPA–SVM; **f** ground truth segmentations

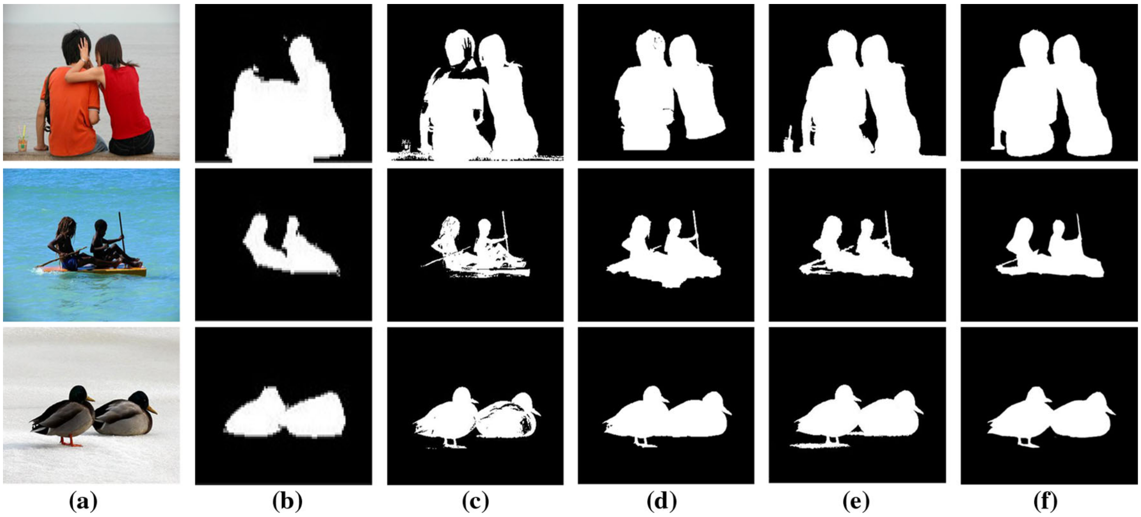

Fig. 5 Comparison of salient object segmentation. **a** Original images; **b** segmentation result using Ncut; **c** segmentation results using Kmeans; **d** segmentation results using RCC; **e** segmentation results of PPA–SVM; **f** ground truth segmentations

good segmentation results in all test images. Additionally, for some very small salient object as in the first image, the bottle besides the people can be segmented completely. Only for the third image, the shaded area under the bird is false-segmented due to its color is similar to the salient object.

In summary, for *9* test images listed above, most salient objects are correctly detected and extracted, the segmentation results of PPA–SVM are closest to the ground truth segmentation in most cases. Meanwhile, some more detailed texture and color information about salient object interior are also considered. What's more, PPA–SVM method can segment some smaller salient objects in some images.

Table 1 presents the quantitative evaluations (ER, PRI, VI, GCE) for four different image segmentation algorithms, in which the black values indicate the best results. And ↑ means the larger the metric, the better the segmentation result, and vice versa. It can be seen that PPA–SVM algorithm are better than the other three algorithms in most cases. The mean ER, PRI, VI and GCE values of *9* test images are 2.24 %, 0.95, 0.27 and 0.037, respectively.

And the CPU time comparisons of four methods are shown in Fig. 6, which indicates that PPA–SVM method has an advantage in segmentation speed. The CPU time of PPA–SVM is significantly less than the time required by the algorithms of Ncut and RCC. Computational

Table 1 Comparisons of 4 evaluation metrics for four methods of 9 test images

	ER(%) ↓				PRI ↑				VI ↓				GCE ↓			
	Kmeans	Ncut	RCC	Ours	Kmeans	Ncut	RCC	Ours	Kmeans	Ncut	RCC	Ours	Kmeans	Ncut	RCC	Ours
img1	4.3	5.3	5.6	**3.4**	0.92	0.90	0.89	**0.95**	0.39	0.52	0.54	**0.39**	0.07	0.09	0.10	**0.03**
img2	2.5	6.1	0.3	**0.2**	0.95	0.89	0.91	**0.96**	0.29	0.57	0.06	**0.04**	0.05	0.10	0.02	**0.01**
img3	1.8	5.1	0.9	**0.4**	0.96	0.91	0.92	**0.97**	0.22	0.50	0.21	**0.14**	0.03	0.08	0.03	**0.02**
img4	14.7	23.3	**0.3**	4.1	0.75	0.64	**0.99**	0.92	1.07	1.20	**0.05**	0.46	0.21	0.29	**0.01**	0.07
img5	8.0	9.6	11.7	**3.1**	0.85	0.83	0.80	**0.95**	0.61	0.83	0.69	**0.29**	0.10	0.15	0.10	**0.04**
img6	2.6	4.6	2.9	**2.1**	0.95	0.91	0.94	**0.96**	0.33	0.50	0.30	**0.28**	0.05	0.08	0.05	**0.04**
img7	9.7	15.2	3.5	**1.5**	0.82	0.74	0.94	**0.97**	0.88	1.18	0.29	**0.20**	0.17	0.24	0.03	**0.01**
img8	3.7	5.8	3.4	**3.2**	0.93	0.89	0.92	**0.94**	0.40	0.56	0.36	**0.34**	0.06	0.09	0.06	**0.05**
img9	2.5	4.9	**0.9**	2.1	0.95	0.90	**0.98**	0.97	0.30	0.49	**0.13**	0.18	0.05	0.08	**0.01**	0.02

Bold values indicate the best results

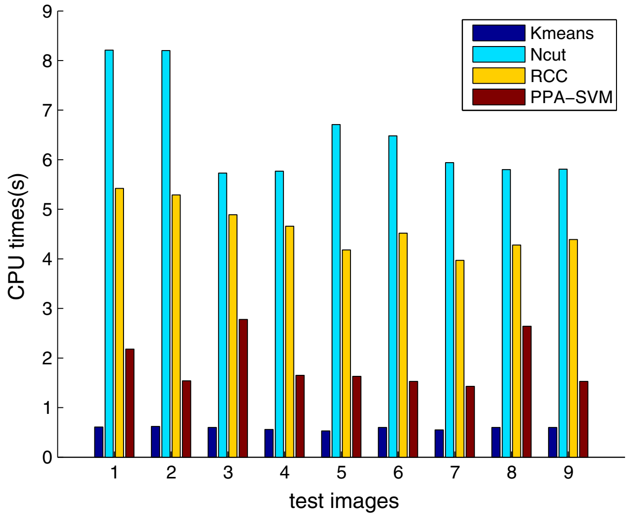

Fig. 6 CPU times required to segment salient objects in 9 test images using Kmeans, Ncut, RCC and PPA–SVM method. Algorithms were implemented in a Core 2.6-GHz computer with 2-GB RAM

complexity of Kmeans and PPA–SVM are, respectively, $\mathcal{O}(Nk)$ and $\mathcal{O}(n)$, where N is the total number of image pixel, k is the number of cluster, and n is the number of training pixels in PPA–SVM. In most cases, n is much smaller than N, but PPA–SVM requires salient region extraction and histogram peak selection, so compared with Kmeans method, PPA–SVM method is little slower, but it can produce superior quality-segmented results. And because each pixel is taken as a node of a graph in Ncut method, its computational complexity is $\mathcal{O}(N^{3/2})$, so the computation speed of Ncut is the largest in four segmentation methods. While for RCC method, whose computational complexity is $\mathcal{O}(N)$, but Grabcut is iteratively applied to refine the segmentation result initially obtained by thresholding the saliency map, so some extra time is needed. For PPA–SVM method, CPU times used for segmenting salient object in 9 test images are no more than 3 seconds, which demonstrates that the proposed PPA–SVM algorithm can be applied in real-time applications.

Some other visual comparisons of segmentation results obtained by these four segmentation methods are shown in Fig. 7. In these test images, some salient objects are with black shadows (like the first and the second images) that will affect the segment results, and some image background color distribution is not homogeneous (like the third image). It can be seen from Fig. 7, the PPA–SVM method outperforms other three methods in most cases. The salient objects in test images can be segmented more accurately from background, even some texture information in salient object can be extracted in some images. Experiment results on the whole dataset support this conclusion but not listed with the limit of paper length. Comparisons of average values of 5 metrics (ER, PRI, VI, GCE, and CPU times) on the whole dataset are almost consistent with the visual comparisons, as shown in Table 2. Obviously, PPA–SVM algorithm is better than other algorithms in 4 evaluation metrics (ER, PRI, VI and GCE), but the average CPU time of PPA–SVM is slightly inferior to that of Kmeans. Above experiment results demonstrate that PPA–SVM method can effectively detect and segment salient object in relatively less times.

Finally, similar to Refs. [13, 16], average precision, recall, and F-Measure are compared over the entire dataset too. F-Measure is defined as:

$$F_\beta = \frac{(1+\beta^2)Precision \times Recall}{\beta^2 \times Precision + Recall}. \qquad (19)$$

We use $\beta^2 = 0.3$ as Refs. [6, 16] to measure precision more than recall. Compared with the state-of-the-art results

Fig. 7 Comparison of other segmentation results with two different segmentation methods and RCC saliency detection method. **a** Original images; **b** segmentation results of Ncut; **c** segmentation results of Kmeans; **d** segmentation results of RCC; **e** segmentation results of PPA–SVM; **f** ground truth segmentations

▶

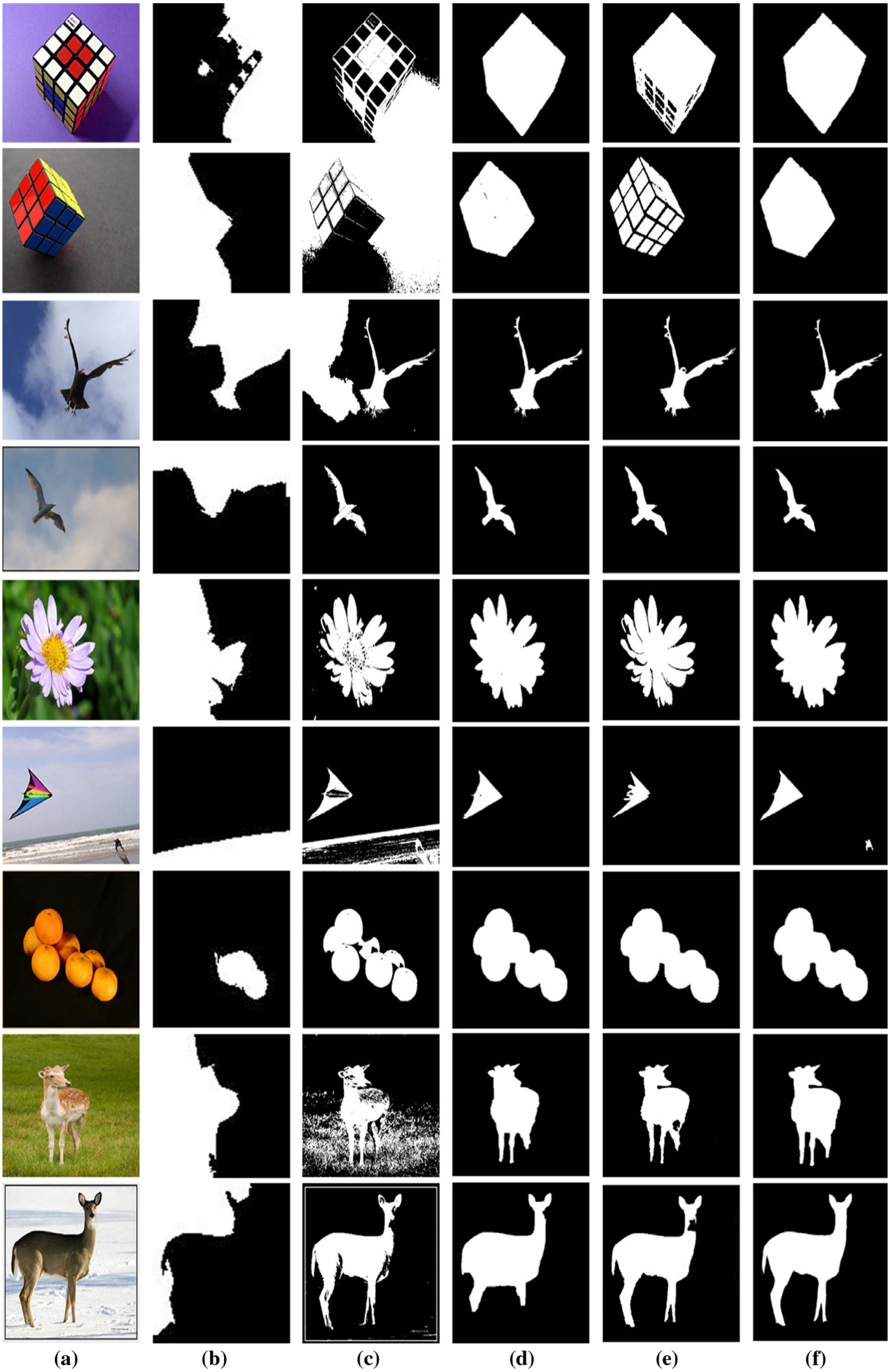
(a)
(b)
(c)
(d)
(e)
(f)

Table 2 Comparisons of 5 evaluation metrics for four methods on the whole dataset

	ER(%)↓	PRI ↑	VI↓	GCE↓	CPU↓
Kmeans	12.1	0.813	0.744	0.126	**1.98**
Ncut	8.87	0.847	0.706	0.132	8.52
RCC	4.37	0.905	0.424	0.09	4.62
PPA–SVM	**4.26**	**0.916**	**0.408**	**0.07**	2.38

Table 3 Comparative results in classification accuracy, the number of support vectors and CPU time of different number of training pixels

Value of T_h	0	10	20	50	100
Classification accuracy (%)	96.6	95.8	96.8	96.9	96.9
Number of support vectors	12	36	59	138	246
CPU time	2.7	4.2	5.8	8.3	12.5

on this dataset by Achanta et al. (precision = 75 %, recall = 83 %) [16] and RCC (precision = 90 %, recall = 90 %) [6], better accuracy (precision = 92 %, recall = 93 %) can be achieved for PPA–SVM method.

Overall, the proposed PPA–SVM is an automatic image segmentation method based on SVM classifier, so the commonly used metrics for classifier performance such as classification accuracy, generalization ability should also be taken into consideration. Although this is not the main focus of this paper, there are many new refinement techniques by maximizing the uncertainty or combining multiple classifiers have been proposed to improve the generalization capability of the learning system [47–49]. And it is well known the classification performance of SVM classifier heavily depends on the number and distribution of training examples. Therefore, in order to evaluate the property of PPA–SVM from the perspective of classification performance, a series of the number of training pixels are selected according to the local homogeneity threshold T_h to train the PPA–SVM.

Table 3 gives the comparative results in classification accuracy, the number of support vectors, CPU time of different number of training pixels based on different local homogeneity thresholds. It can be seen that although the number of support vectors, the training time and the segmenting time are changed to varying degrees as the number of training pixels increases, the classification accuracy is slightly changed. When the threshold value T_h is very high, it means that the pixels in training sets TS_p and TS_n are all selected as training pixels, which leads to more training time. Conversely, when the threshold T_h is set to zero, only those with the highest homogeneity are selected as training pixels. Thus, the training time is rapidly reduced, but the classification accuracy has little effect. Therefore, the homogeneity threshold for selecting reprehensive pixels in local region is effective in terms of segmenting speed and accuracy, especially for real-time image segmentation tasks.

From above experiments and analysis, the performance of PPA–SVM is much superior to that of Kmeans, Ncut, and RCC. Although RCC method could achieve better performance in some cases, it first needs a pre-segmentation results produced by a graph-based method, and later Grabcut is applied to refine the segmentation result initially obtained by thresholding the saliency map. In comparison, PPA–SVM method does not rely on any prior information or pre-segmentation results. In other words, PPA–SVM is a pure bottom-up low-level feature analysis processing. The better performance may come from the excellent performance of SVM classifier and principal pixel analysis.

4 Conclusion

In this paper, a novel automatic salient object segmentation approach based on principal pixel analysis and SVM classifier is proposed. The advantages of the proposed PPA–SVM method can be concluded as: (1) Global saliency information, spatial location, and local color features are all considered, and not any prior knowledge about the salient object such as color, texture, intensity is needed. (2) The whole salient object with homogeneous features can be extracted. (3) The salient object can be segmented automatically and adaptively without human intervention. (4) Saliency detection and principal pixel analysis can help to select the SVM training samples automatically, which avoids the problems caused by manual intervention and random selection. Compared with some typical segmentation methods, PPA–SVM method can provide better segmentation results with less time simultaneously. In particular, PPA–SVM can obtain more detailed visual information about salient object, which will provide more cues for further object recognition and understanding tasks. It should be noted that, PPA–SVM is effective for images with significant salient object, and it may be affected for images with more complex color and texture. At present, PPA–SVM only takes into account some low-level information such as color, texture and spatial location. If some higher-level information that based on semantic or task-orientated is added, PPA–SVM will produce more effective segmentation for greater variety of images. This will be our further works.

Acknowledgments The work described in this paper was partially supported by the National Natural Science Foundation of China (No. 61273291), Research Project Supported by Shanxi Scholarship

Council of China (No. 2012-008), Scientific and Technological Project of Shanxi Province (No. 20120321027-01), Open Project Foundation of Information Technology Research Base of Civil Aviation Administration of China (No. CAAC-ITRB-201305).

References

1. Hiremath PS, Jagadeesh P (2008) Content based image retrieval using color boosted salient points and shape features of an image. Int J Image Process 2(1):1–34
2. Guo CL, Zhang LM (2010) A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Trans Image Process 19(1): 185–198
3. Yu Y, Mann GKI, Gosine RG (2010) An object-based visual attention model for robotic applications. IEEE Trans Syst Man Cybern B Cybern 40(5):1398–1412
4. Li H, Ngan KN (2008) Saliency model based face segmentation in head-and-shoulder video sequences. J Vis Commun Image Represent 19(5):320–333
5. Tsotsos JK, Culhane SM, Wai WYK, Lai Y, Davis N, Nuflo F (1995) Modelling visual attention via selective tuning. Artif Intell 78(1–2):507–545
6. Cheng MM, Zhang GX, Mitra NJ, Huang X, Hu SM (2011) Global contrast based salient region detection, CVPR 21–23
7. 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
8. Harel J, Koch C, Perona P (2007) Graph-based visual saliency. Adv Neural Inf Process Syst 19:545–552
9. Lin Y, Fang B, Tang Y (2010) A computational model for saliency maps by using local entropy. Proc Conf AAAI Artif Intell 967–973
10. Walter D, Koch C (2006) Modelling attention to salient protoobject. Neural Netw 19(9):1395–1407
11. Valenti R, Sebe N, Gevers T (2009) Images saliency by isocentric curvedness and color, ICCV 2185–2192
12. Ma YF, Zhang HJ (2003) Contrast-based image attention analysis by using fuzzy growing. International conference on multimedia, pp 374–381
13. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum HY (2011) Learning to de tect a salient object. IEEE Trans Pattern Anal Mach Intell 33(2):353–367
14. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach, CVPR 1–8
15. Guo C, Ma Q, Zhang L (2008) Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform, CVPR 1–8
16. Achanta R, Hemami S, Esgtrada F, Süsstrunk S (2009) Frequency-tuned salient region detection, CVPR 1597–1604
17. Rosin PL (2009) A simple method for detecting salient regions. Pattern Recognit 42(11):2363–2371
18. Luo W, Li H, Liu G, Ngan KN (2011) Global salient information maximization for saliency detection. Sig Process Image Commun 27(3):238–248
19. Yang W, Tang Y, Fang B, Shang Z, Lin Y (2013) Visual saliency detection with center shift. Neurocomputing 103(1):63–74
20. Ouerhani N, Archip N, Hügli H, Erard PJ (2001) Visual attention guided seed selection for color image segmentation. Proceedings of the 9th international conference on computer analysis of image and patterns, lecture notes in computer science, vol 2124, Springer, London, pp 630–637
21. Han J, Ngan KN, Li M, Zhang HJ (2006) Unsupervised extraction of visual attention objects in color images. IEEE Trans Circuits Syst 16(1):141–145
22. Ko BC, Nam JY (2006) Object-of-interest segmentation based on human attention and semantic region clustering. J Opt Soc Am A 23(10):2462–2470
23. Achanta R, Estrada F, Wils P, Süsstrunk S (2008) Salient region detection and segmentation. In: Proceedings of the 6th international conference on computer vision systems, lecture notes in computer science, vol. 5008, Springer, Berlin pp 66–75
24. Donoser M, Urschler M, Hirzer M, Bischof H (2009) Saliency driven total variation segmentation, ICCV 817–824
25. Liu Z, Li W, Shen L, Han Z, Zhang Z (2010) Automatic segmentation of focused objects from images with low depth of field. Pattern Recognit Lett 31(7):572–581
26. Lee CY, Leou JJ, Hsiao HH (2012) Saliency-directed color image segmentation using modified particle swarm optimization. Signal Process 92(1):1–18
27. Liu Z, Shen L, Zhang Z (2011) Unsupervised image segmentation based on analysis of binary partition tree for salient object extraction. Signal Process 91(2):290–299
28. Zhu R, Yao M, Liu YM (2011) A two-level strategy for segmenting center of interest from pictures. Expert Syst Appl 38(3):1748–1759
29. Fu K, Gong C, Yang J, Zhou Y, Gu IYH (2013) Superpixel based color contrast and color distribution driven salient object detection. Signal Process Image Commun 28(10):1448–1463
30. Vapnik VN (1995) The nature of statistical learning theory. Spring, New York
31. Yu Z, Wong HS, Wen G (2011) A modified support vector machine and its application to image segmentation. Image Vis Comput 29(1):29–40
32. Wang XY, Wang T, Bu J (2011) Color iamge segmentation using pixel wise support vector machine classification. Pattern Recognit 44(4):777–787
33. Saha I, Maulik U, Bandyopadhyay S, Plewczynski D (2012) SVMeFC: SVM ensemble fuzzy clustering for satellite image segmentation. IEEE Trans Geosci Remote Sens 9(1):52–55
34. Zhao Q, Hu Y, Cao J (2009) Automatic image segmentation based on saliency maps and Fuzzy SVM, CCWMC 121–124
35. Castleman KR (1996) Digital image processing, second ed. Prentice Hall, New York
36. Wang P, Wang J, Zeng G, Feng J, Zha H, Li S (2012) Salient object detection for searched web images iva global saliency. In CVPR, 3194–3201
37. Brigger P, Casas JR, Pardas M (1996) Morphological operators for image and video compression. IEEE Trans Image Process 5(6):881–898
38. Chen TW, Chen YL, Chien SY (2008) Fast image segmentation based on K-means clustering with histograms in HSV color space. IEEE workshop multimed. Signal Proc pp 322–325
39. Zhang L, Lin FZ, Zhang B A CBIR method based on color-spatial feature, Technical report, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
40. MacQueen JB (1967) Some methods for classification and analysis of multivariate observations. Proceedings of the 5th Berkeley symposium on mathematical statistics and probability 2:281–297
41. Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell 22(8):888–905
42. Rother C, Kolmogorov V, Blake A (2004) "Grabcut"–interactive foreground extraction using iterated graph cuts. ACM Trans Graph 23(3):309–314
43. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2(3):1–27. (http://www.csie.ntu.edu.tw/cjlin/libsvm)
44. Unnikrishnan R, Pantofaru C, Hebert M (2007) Toward objective evaluation of image segmentation algorithms. IEEE Trans Pattern Anal Mach Intell 29(6):929–943

45. Meila M (2005) Comparing clusterings: an axiomatic view. ICML 577–584
46. 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, ICCV 416–425
47. Wang XZ, Dong CR (2009) Improving generalization of fuzzy if-then rules by maximizing fuzzy entropy. IEEE Trans Fuzzy Syst 17(3):556–567
48. Wang XZ, Zhai JH, Lu SX (2008) Induction of multiple fuzzy decision trees based on rough set technique. Inf Sci 178(16): 3188–3202
49. Zhai JH, Xu HY, Wang XZ (2012) Dynamic ensemble extreme learning machine based on sample entropy. Soft Comput 16(9): 1493–1502