http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 7 (7): 992-1000, 2008
ISSN 1812-5638

Salient Region: Presentations of Image Main Contents and its Exaction Algorithms

[1,2]Mingxin Zhang, [1]Zhaogan Lu and [3]Junyi Shen
[1]Department of Information and Communication Engineering, Xi'an Jiaotong University, China
[2]Department of Computer Engineering, Lanzhou Polytechnic College, China
[3]College of Computer Science, Chongqing University, China

Abstract: In order to accurately detect those salient useful regions with flat textures, we present a novel saliency metric coupled with color and texture features and its corresponding salient region extraction methods. Salient regions are usually defined as those regions that could present the main meaningful or semantic contents. However, there are no uniform saliency metrics that could describe the saliency of implicit image regions. Most common metrics take those regions as salient regions, which have many abrupt changes or some unpredictable characteristics and this will fail to detect those salient useful regions with flat textures. In fact, according to human semantic perceptions, color and texture distinctions are the main characteristics that could distinct different regions. In this study, three main colors and multi-resolution Gabor features are respectively used to evaluate the corresponding saliency values of implicit regions in one image. For each region, its saliency value is actually to evaluate the total sum of its Euclidean distances for other regions in the color and texture spaces. A special synthesized image and several practical images with main salient regions are used to evaluate the performance of the proposed saliency metric and other several common metrics, i.e., scale saliency, wavelet transform modulus maxima point density and important index based metrics. Experiment results verified that the proposed saliency metric could achieve more robust performance than those common saliency metrics.

Key words: Salient regions, color and texture features, image segmentation, saliency metric

INTRODUCTION

The research data (Jun-Wei *et al*., 2000; Feng *et al*., 2004) of content-based image retrieval (CBIR) over the last few years has shown that retrieving images through matching images solely on that basis of global similarities is often too crude to produce satisfactory results. On the other hand, semantic object-based image retrieval (Celebi and Alpkocak, 2006) is still far too rudimentary and fragile to produce reliable results. Intermediate-level processing (Pappas *et al*., 2007) between high and low-level processing for content-based image retrieval is required. Therefore, it is necessary to identify the perceptually salient and semantically meaningful regions (Jianping *et al*., 2005) in images. However, it is difficult to isolate the meaningful region of interest from the scene without a priori knowledge (Dadir and Brady, 2001). In a common case, the regions with many abrupt changes or some unpredictable characteristics (Ling and Michael, 2006; Zhang *et al*., 2006) often attract the human's attention, are considered as the salient regions of images in this study. Thus, salient regions of one image are those regions that could present the main contents of the image, which were detected according to local features as such as colors, textures and shapes. Moreover, we believe that these salient regions are potentially more effective for image indexing, retrieval and classification.

According to the above definition of salient regions, Kadir and Brady (2001) proposed a salient region detection method Scale Saliency and its improved versions are given by Shao *et al*. (2007). Although this method could detect those regions with rich information in terms of information entropy, this method must be evaluated for every pixel with different size vicinities. As an alternative approaches, salient points-based approaches are presented by Dai *et al*. (2006) and Chul *et al*. (2004), where the salient regions were detected according to the density of salient points in each local segmented region. However, there exists fatal drawback for both the Scale Saliency method and these salient points-based approaches, i.e., as they could not detect those meaningful regions with flat textures and few salient points. An important index was used as the metric of region saliency and the sizes of implicit image regions could also be reflected in its definition (Yu-Hsin *et al*., 2006).

Corresponding Author: Mingxin Zhang, Department of Information and Communication Engineering, Xi'an Jiaotong University, China

However, as the complexity of practical image analysis, ones could not deal with every region for different sizes and it's very important to exclude those implicit regions smaller than the specified size. Furthermore, the salient regions should be larger than the given region sizes and with rich semantic information to human being. So, we proposed a novel salient region detection method, which could take account of the implicit region sizes and their saliency relative to other regions. Firstly, all the regions smaller than the given sizes are excluded from the candidate regions, then the most salient will be picked out after their saliency relative to other candidate regions were evaluated. The saliency could be achieved according to their feature vectors and these feature vector consist of main colors and texture features, where the texture features are presented by the means and variances of Gabor coefficients at several different frequencies and orientations as given by Kamarainen *et al*. (2006) and Arivazhagan *et al*. (2006) and three main colors in every region are extracted by the mean-shift cluster method (Comaniciu and Meer, 2002). Thus, the proposed scheme could pick out the most salient regions in terms of people perception and region size requirements than the approaches given by Dadir and Brady (2001), Dai *et al*. (2006), Shao *et al*. (2007), Chul *et al*. (2004) and Yu-Hsin *et al*. (2006).

SALIENCY OVERVIEW

As we know, one image could be presented completely by one or several specific salient regions. So, it's pivotal to figure out the definition of salient regions that are consistent with peoples' perception. The saliency definitions as showed in those attempts (Dadir and Brady, 2001; Ling and Michael, 2006; Dai *et al*., 2006; Shao *et al*., 2007; Chul *et al*., 2004; Yu-Hsin *et al*., 2006), are actually some aspects of salient regions for human perception, as they ignore the adaptation of human perception under different scenarios or objective targets. Here, in order to disclose the drawbacks of current common saliency, we list three typical saliency definitions presented in literatures (Ling and Michael, 2006; Dai *et al*., 2006; Yu-Hsin *et al*., 2006), respectively, where the details of these methods could be found out.

Scale saliency: Ling and Michael (2006) proposed a salient region detection method they call Scale Saliency. The algorithm deems salient those regions exhibiting unpredictable characteristics simultaneously in some feature-space and over scale. It is a product of two terms, both a function of the Probability Density Function (PDF) of local image attributes (e.g., intensity, color) at multiple scales.

The first, H_D, is Shannon Entropy and measures feature space unpredictability, the second, W_D, measures inter-scale unpredictability. Extrema in H_D are used as the basis for scale selection as the basis for scale selection. In the discrete case, Scale Saliency is defined as:

$$Y_D(s_p, X) = H_D(s_p, X) \times W_D(s_p, X) \quad (1)$$

where, entropy H_D is defined by:

$$H_D(s, X) = -\sum_{d \in D} p_{d,s,X} \log_2 p_{d,s,X} \quad (2)$$

where, $p_{d,s,x}$ is the probability as a function of scale s, position X and descriptor value d which takes values in D, the set of all descriptor values. The inter-scale saliency measure, W_D, is defined by:

$$W_D(s, X) = \frac{s^2}{2s-1} \sum_{d \in D} |p_{d,s,X} - p_{d,s-1,X}| \quad (3)$$

The set of scales s_p, at which entropy peaks, is defined by:

$$s_p = \{s \mid H_D(s-1, X) < H_D(s, X) < H_D(s+1, X)\} \quad (4)$$

In principle, the above algorithm possesses a number of attractive properties: robustness to similarity transformations, e.g., planar rotation, spatial scaling, translation and intensity shifts and scaling. However, in practice, the performance of the algorithm under geometric and photometric transforms falls short of the theoretical expectations. Furthermore, the saliency in (1) is defined for every pixel with multiple different scale vicinities and this means it must be done for every pixel of one image. It also could not detect those semantic salient regions with flat textures, as the semantic salient regions don't need be consistent with abrupt changes or unpredictable characteristics.

Wavelet transform modulus maxima point density: Mallat and Hwang (1992) have proposed the use of wavelet transform modulus maxima (WTMM) for characterization of regularity of signals. By examining the wavelet modulus maxima, they could measure the local Lipschitz exponents of a signal. This turned out to be an effective way to locate the singularity or edges of an image. The wavelet modulus maxima can represent the sharply focused edges and the inner texture details of the interest regions in image. Therefore, the wavelet maxima point density of the image is in proportion to the visual complexity of the whole image and so the saliency regions are. The extraction is processed based on the results of wavelet modulus maxima (WMM) edge detection and mean shift color region segmentation. Because regions having abundant various details often attract humans' attention, but not does those invariable background. We classify the color regions to salient interest region or background according to their WMM point densities.

Suppose an image is divided into N color regions denoted by $\{R_i\}_{i=1...N}$ and the centers of color regions by $\{C_i\}_{i=1...N}$. Let $\{W_k\}_{k=1...P}$ be the WMM points translated from the image. The details of the classification algorithm are provided in the list below as:

- For each WMM point W_k, Compute the Euclidean distances d_k from W_k to each color region center C_i, where k = 1... P, i = 1…N. The point W_k is attached to the region C_l, where $d_l^k = \min_{0<i\le N}(d_i^k)$, $0 < l \le N$.
- For each color region R_i, compute its WMM points count E_i included, its region pixels count $Area_i$ and its boundary pixels count $Blen_i$, i = 1...N.
- For each color region R_i , if $E_i/Area_i > th_1$ or $E_i/Area_i > th_2$, then $R_j \in R_{salient}$, else $R_j \in R_{background}$ where th_1 and th_2 are two thresholds.

The density of wavelet modulus maxima points actually reveals the degree of abrupt changes in one region and in some instances the salient regions is consistent with those abrupt change regions. But, for those salient object regions with flat textures, it will fail to find out these semantic salient regions.

Important index: Both the definitions of scale saliency and WTMM-based salient regions assume that the salient regions are those with abrupt changes or unpredictable characteristics. They can be considered as the reflection of region information entropy, but ignore the implicit region sizes. Thus, another region size-based saliency is given by Ling and Michael (2006). A salient region should be compact, complete and significant enough and neither a small region nor a fragmentary region can be important. Thus, the important index of region is defined as follows:

$$I(R_j^i) = \left(\frac{N_{R_j^i}}{\sum_{j=1}^{m_i} N_{R_j^i}} \right) \times \left(\frac{N_{R_j^i}}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} N_{R_j^i}} \right) \tag{5}$$

where, R_j^i is the region with color label i and region index j, $I(R_j^i)$ and $N_{R_j^i}$ denote the important index and the pixel number of region R_j^i, respectively. $\sum_{j=1}^{m_i} N_{R_j^i}$ is the total number of pixels of all regions with color label i, while $\sum_{i=1}^{n}\sum_{j=1}^{m_i} N_{R_j^i}$ is the total number of pixels of an image.

Although the definition of salient regions in (5) is simple, it does not consider the comparisons among different regions in terms of colors, textures and information entropy. It maybe fails to detect some regions with salient features but small sizes. Thus, it is necessary to give a novel metric of saliency that can be used to figure out the salient regions on the base of the comparisons among different regions in terms of sizes, colors, textures and information entropy.

THE PROPOSED SCHEME

Aiming at the drawbacks of salient region detection methods given in last section, we propose a new saliency metric which can describe the salient regions more accurately than those approaches. As pointed out by Yu-Hsin *et al*. (2006), one salient region should be compact, complete and significant enough and neither a small region nor a fragmentary region can be one meaningful region. Furthermore, in terms of colors and textures, the salient regions should also have distinguishing features from their neighbor regions or other regions. Provided that each region is characterized by its feature vectors, they are the syntheses of color and texture features and the new saliency metric could be outlined in the follows:

The color features consist of three main colors that can be achieved by color clusters of pixels in the region, while the texture features are described by the means and variances of Gabor filter bank at different frequencies and orientations. Thus, the color features of the i-th region are denoted as $Fc(i) = \{[r(k,i),g(k,i),b(k,i)] | k = 1,2,3\}$, i = 1,2,…, N, where N is the total number of region and (r_j,g_j,b_j) is the j-th main color in current regions, respectively. If P scales and Q orientations are considered in the Gabor bank, the textures of the i-th region are denoted as $F_t(i) = \{\mu_{m,n}(i), \sigma_{m,n}(i) \mid m = 1,2,\ldots,P;\ n = 1,2,\ldots,Q\}$, where $\mu_{m,n}(i)$ and $\sigma_{m,n}(i)$ are the mean and variance of the Gabor filter coefficients for i-th region at m-th frequency and n-th orientation.

According to the feature vectors of all the regions, the saliency of one region is given as the sum of its Euclidean distance from all the other regions, i.e.,

$$s(i) = \sum_{j=1, j\ne i}^{N} \frac{1}{2}\left[\frac{d_c(i,j)}{\max_{i,j}(d_c(i,j))} + \frac{d_t(i,j)}{\max_{i,j}(d_t(i,j))} \right] \tag{6}$$

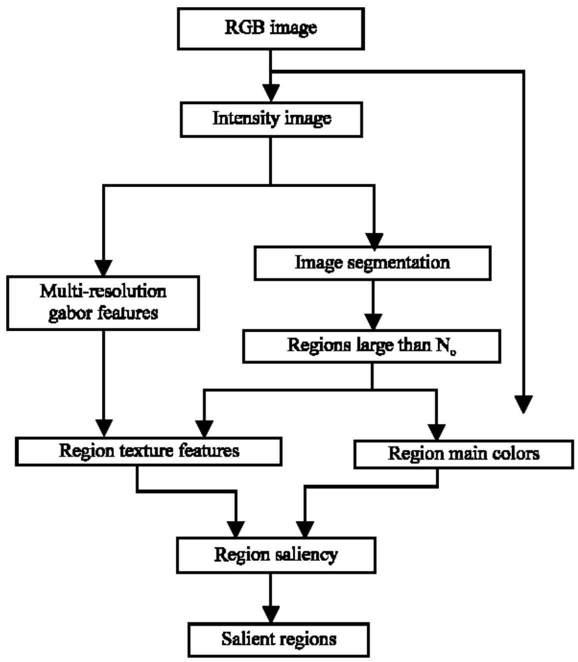

Fig. 1: The chart of salient region detection according to the proposed scheme

where, $d_c(i,j)$ and $d_t(i,j)$ are the Euclidean distances between the i-th and j-th region for color and texture features respectively and the two maximum terms are used to normalize the two distances into interval [0,1]. Meanwhile, they can be computed as:

$$d_c(i,j)=\sqrt{\sum_{k=1}^{3}\left[r_k(i)-r_k(j)\right]^2+\left[g_k(i)-g_k(j)\right]^2+\left[b_k(i)-b_k(j)\right]^2} \tag{7}$$

$$d_t(i,j)=\sqrt{\sum_{m=1}^{P}\sum_{n=1}^{Q}\left[\mu_{m,n}(i)-\mu_{m,n}(j)\right]^2+\left[\sigma_{m,n}(i)-\sigma_{m,n}(j)\right]^2} \tag{8}$$

Then, the set of salient regions can be selected as:

$$i_{salient}=\{i \mid s(i)\geq s(j), j=1,2,\ldots,N\} \tag{9}$$

In fact, more than one salient region should be found to meet for the requirements of image analysis, so the set of salient regions are selected as the first several regions according to their saliency values or those regions whose saliency values are greater than one given special threshold. Meanwhile, in order to reduce the computation complexity of saliency values, some small or fragmentary regions must be excluded from the candidate regions, as these regions are not considered as meaningful regions for incoming image processing. The chart of salient region detection is shown in Fig. 1, where N_o is the given minimum region size that one salient region should be.

RESULTS AND DISCUSSION

In order to test the performance of the proposed salient region detection scheme, one synthesized image was constructed with several different balls. Thus, the image has simple regions and could be accurately segmented by most image segmentation methods. There are three balls with distinct colors as showed in Fig. 2 (1-1), whereas the orange ball has no sticks. According to human perception, the orange ball should have the largest saliency values among these small balls and the blue-black and amaranthine balls have inferior saliency respectively. Then, we compare this semantic perception results with that of the proposed scheme to examine its performance. The WTMM point density, scale saliency and important index -based schemes are also conducted to verify the test results of the proposed scheme.

According the saliency value sizes, seven most salient regions are displayed for each salient region detection scheme. As showed in Fig. 2, the test results of proposed scheme are the regions showed by the regions in Fig. 2 (2-[1-7]), while the regions in Fig. 2 (3-[1,7], 4-[1-7],5-[1-7]) are given for the scale saliency, WTMM and important index-based schemes, respectively. The results for our proposed scheme are consistent with that of human semantic perception, while other schemes could not achieve these expectation results. The scheme based important index only selected those regions according to their region sizes, while the WTMM-based scheme selected those regions according to the sizes of the WTMM point density sizes. However, for the salient regions with flat textures, i.e., small WTMM point density, the WTMM-based scheme could detection these saliency regions. The largest region is figured out by the scale saliency-based scheme, as its local information entropy wins its emergence. Its ignorance of color and texture distinction leads to the unexpected results. Their performance differences also could be checked out in the results of other three typical images with different salient regions, as showed in Fig. 3-5, respectively, where three most salient regions were given for each detection schemes. Our proposed scheme could achieve the most salient regions consist with the human semantic perception, which measures the saliency values in terms of color and texture distinctions and has better performance than other current common saliency metrics.

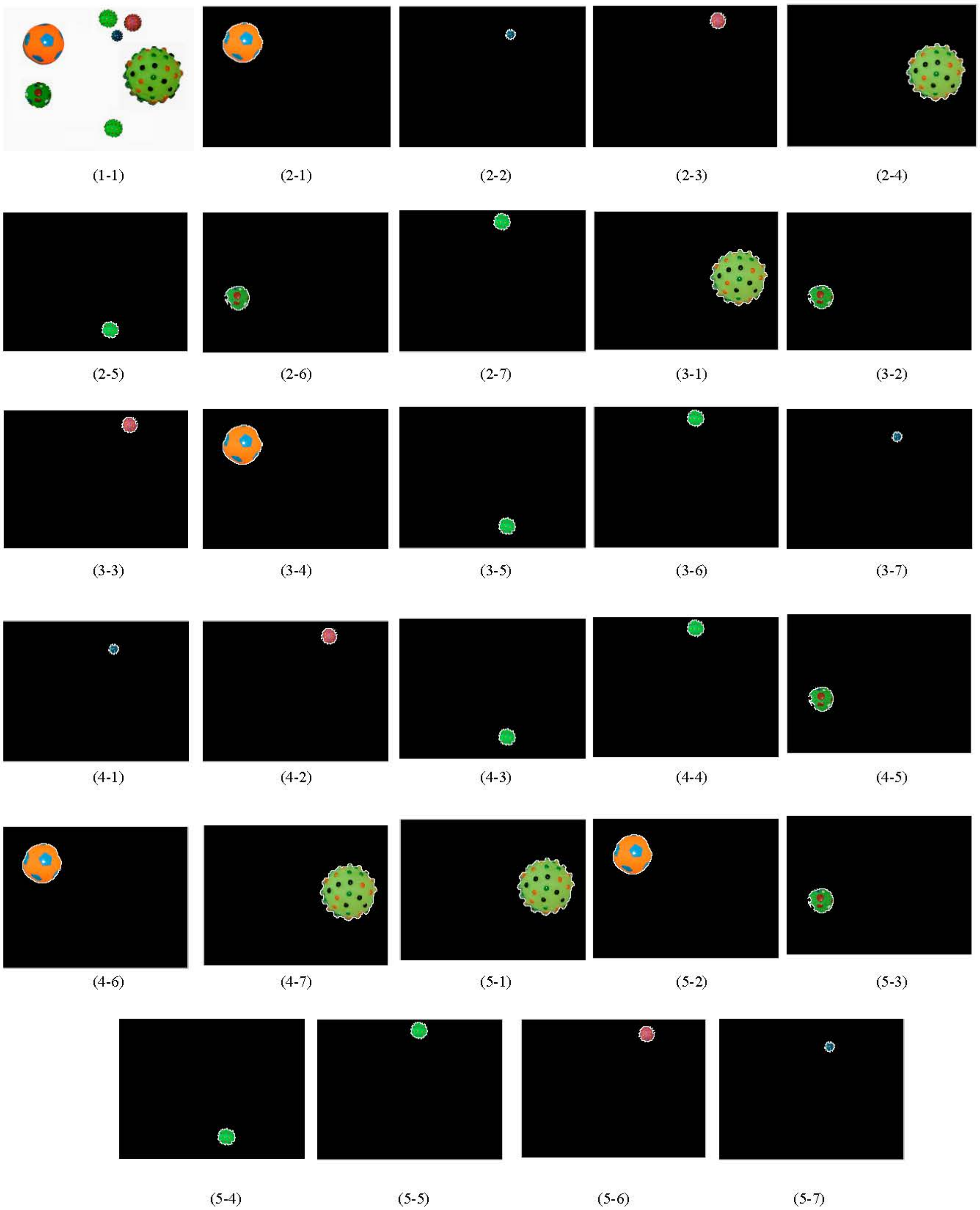

Fig. 2: The synthesized images and the salient regions for the proposed scheme, the scale saliency, WTMM and important index-based scheme respectively, which are given orderly according to the saliency value sizes

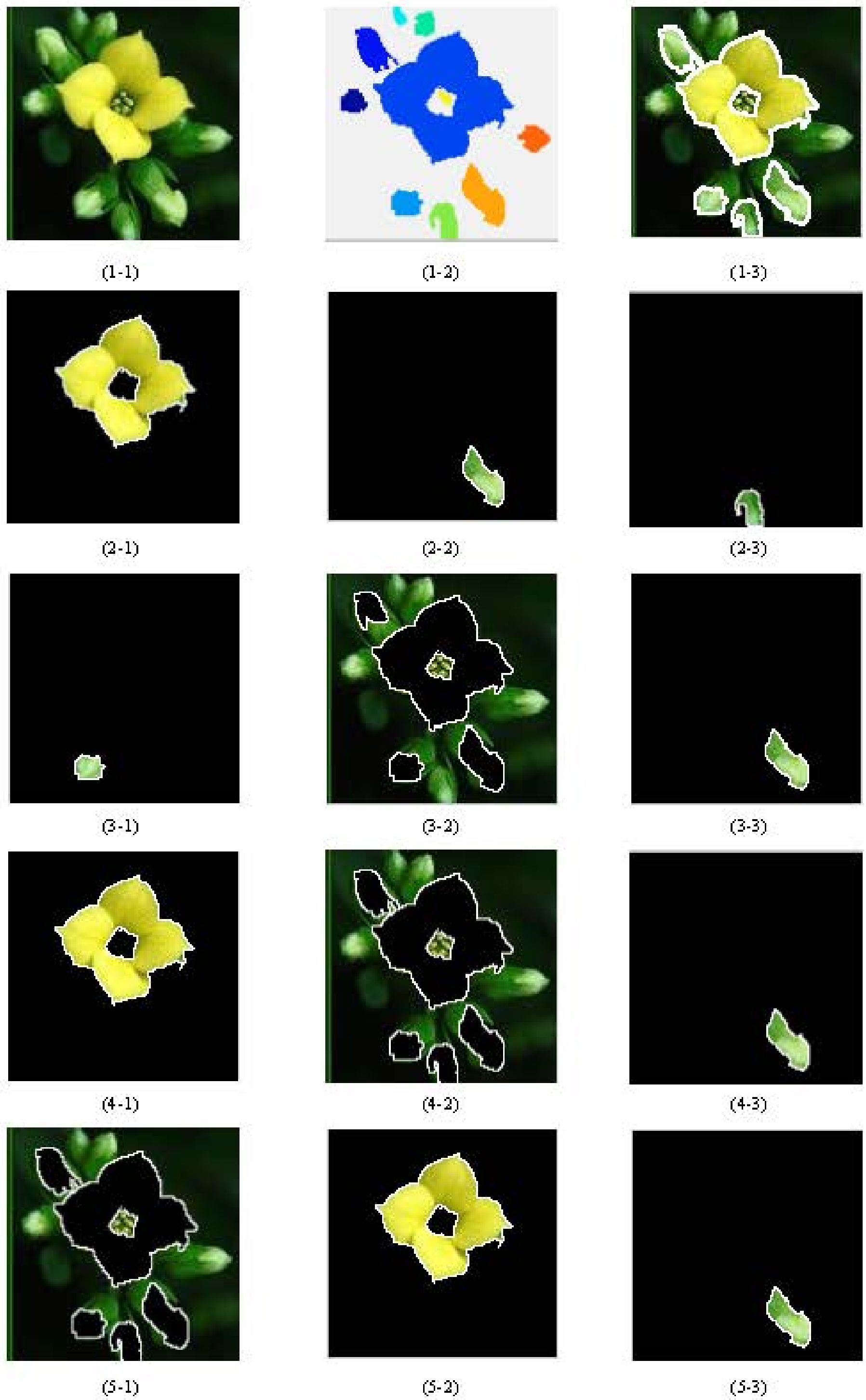

Fig. 3: The determined salient regions for different detection schemes, where (1-1), (1-2), (1-3) are the original image, the segmented regions and the segmented images respectively and (2-3), (3-3), (4-3) and (5-3) are the most salient regions detected by our proposed scheme, the WTMM, scale saliency and important index-based schemes, respectively

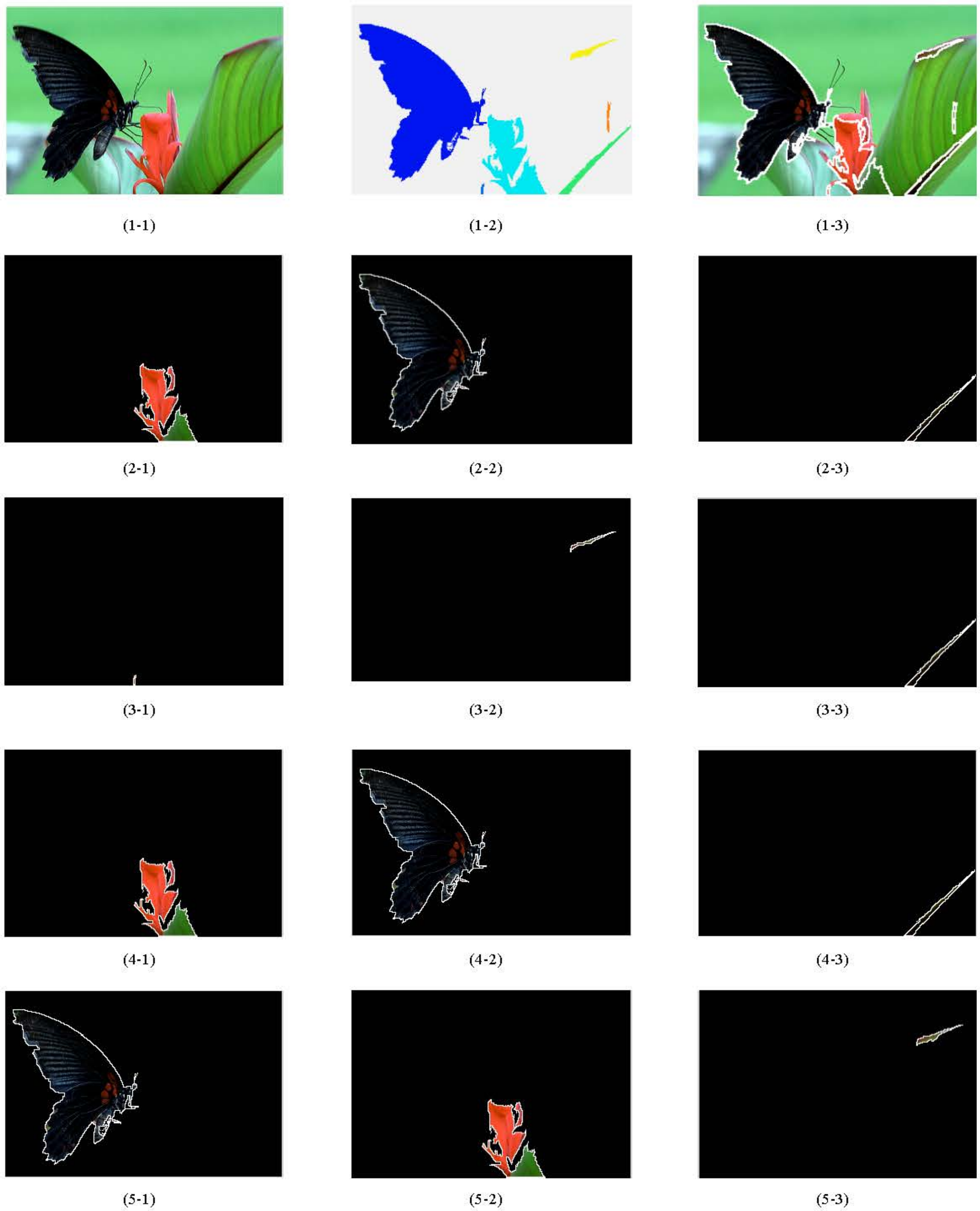

Fig. 4: The determined salient regions for different detection schemes, where (1-1), (1-2), (1-3) are the original image, the segmented regions and the segmented images respectively and (2-3), (3-3), (4-3) and (5-3) are the most salient regions detected by our proposed scheme, the WTMM, scale saliency and important index-based schemes, respectively

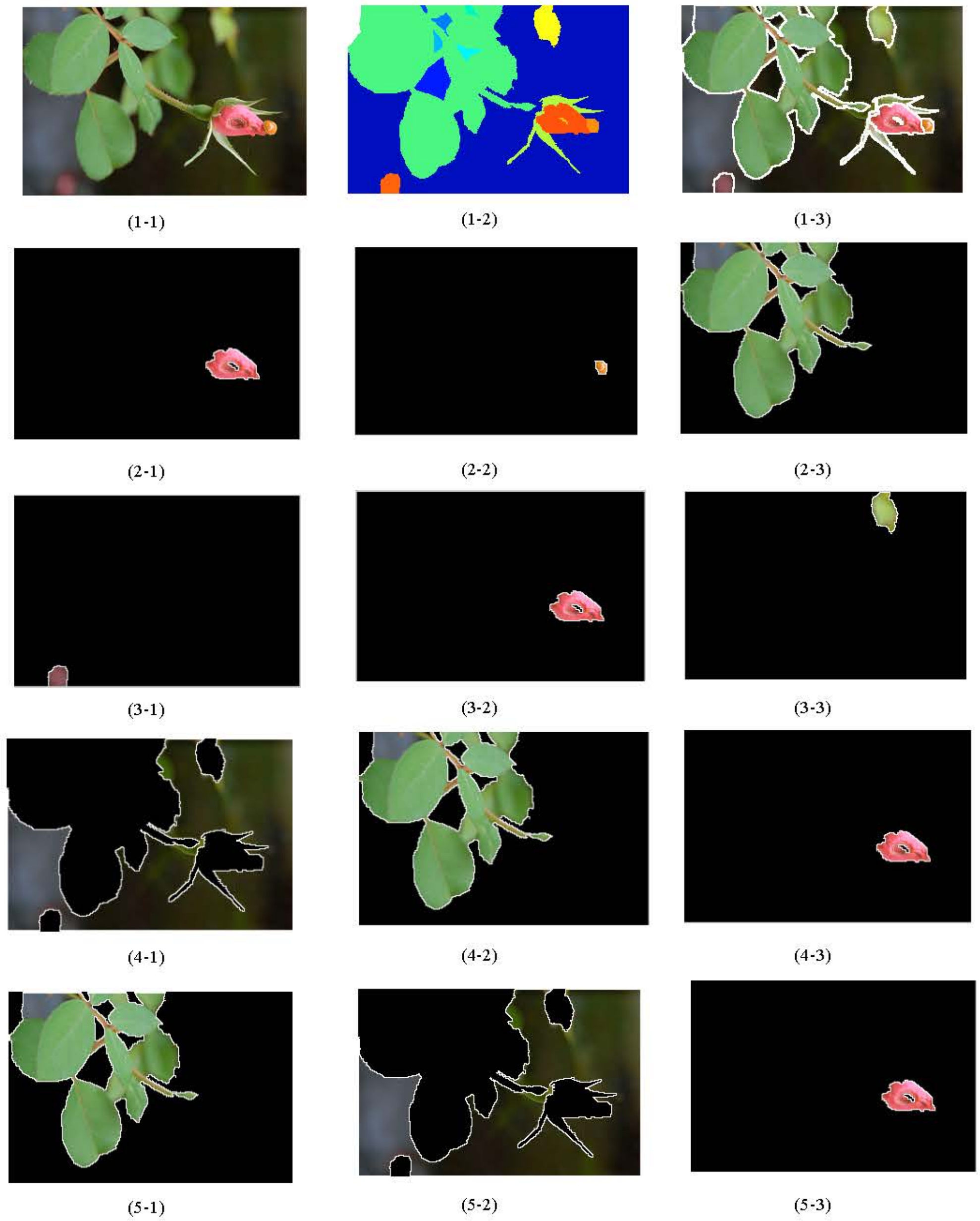

Fig. 5: The determined salient regions for different detection schemes, where (1-1), (1-2), (1-3) are the original image, the segmented regions and the segmented images respectively and (2-3), (3-3), (4-3) and (5-3) are the most salient regions detected by our proposed scheme, the WTMM, scale saliency and important index-based schemes, respectively

CONCLUSIONS

In this study, several common metrics for salient regions are presented, but these metrics could not achieve the consistence to human semantic perceptions. Thus, we propose a novel saliency metric, which can measure the saliency sizes of different regions in context of implicit segmentation regions. The color and texture distinctions are considered to evaluate the saliency metric, as they are the main characteristics that could distinct different regions. A special synthesized image and several practical images with main salient regions are used to evaluate the performance of the proposed saliency metric and other several common metrics. Experiment results verified the better improvements than other common saliency metrics.

REFERENCES

Arivazhagan, S., L. Ganesan and S.P. Padam, 2006. Texture classification using Gabor wavelets based rotation invariant features. Patt. Recog. Lett., 27: 1976-1982.

Celebi, E. and A. Alpkocak, 2006. Semantic image retrieval and auto-annotation by converting keyword space to image space. Proceedings of 2006 12th International Multi-Media Modeling Conference (MMM 2006), Beijing, China, January 2006, Piscataway, NJ, USA, pp: 8-11.

Chul, K.B., K.S. Yeong and B. Hyeran, 2004. SVM-based salient region(s) extraction method for image retrieval. Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), August 23-26, Cambridge UK, pp: 977-980.

Comaniciu, D. and P. Meer, 2002. Mean shift: A robust approach toward feature space analysis. IEEE Trans. Patt. Anal. Machine Intel., 24: 603-619.

Dadir, T. and M. Brady, 2001. Scale, saliency and image description. Int. J. Comput. Vision, 45: 83-105.

Dai, Z.K., L. HanQing and D. MiYi, 2006. Automatic salient regions of interest extraction based on edge and region integration. Proceedings of 2006 IEEE International Symposium on Industrial Electronics (ISIE 2006), July 2006, Montreal Canada, pp: 620-623.

Feng, J., L. Mingjing and Z. Hong-Jiang, 2004. An efficient and effective region-based image retrieval framework. IEEE Trans. Image Process., 13: 699-709.

Jianping, F., G. Yuli and L. Hangzai, 2005. Statistical modeling and conceptualization of natural images. Patt. Recog., 38: 865-885.

Jun-Wei, H., W.E.L. Grimson and C. Cheng-Chin, 2000. Region-based image retrieval. Proceedings of the 2000 International Conference on Image Processing (ICIP 2000), September 10-13, Vancouver, BC, Canada, pp: 77-80.

Kamarainen, J.K., V. Kyrki and H. Kalviainen, 2006. Invariance properties of Gabor filter-based features-overview and applications. IEEE Trans. Image Process., 15: 1088-1099.

Ling, S. and B. Michael, 2006. Invariant salient regions based image retrieval under viewpoint and illumination variations. J. Visual Commun. Image Representation, 17: 1256-1272.

Mallat, S. and W.L. Hwang, 1992. Singularity detection and processing with wavelets. IEEE. Trans. Inform. Theor., 38: 617-643.

Pappas, T.N., C. Junqing and D. Depalov, 2007. Perceptually based techniques for image segmentation and semantic classification. IEEE Commun. Mag., 45: 44-51.

Shao, L., T. Kadir and M. Brady, 2007. Geometric and photometric invariant distinctive regions detection. Inform. Sci., 177: 1088-1122.

Yu-Hsin, K., C. Shih-Ting and K.C. Ming, 2006. A novel unsupervised salient region segmentation for color images. Proceedings of 1st International Conference on Innovative Computing, Information and Control (ICICIC'06), 30 August- 1 September, Beijing, China, pp: 96-99.