Pattern Recognition Letters 34 (2013) 34–41

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Multi-spectral saliency detection

Qi Wang [a,b], Pingkun Yan [a,*], Yuan Yuan [a], Xuelong Li [a]

[a] Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, PR China
[b] School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, Shaanxi, PR China

ARTICLE INFO

Article history:
Available online 15 June 2012

Keywords:
Saliency
Near-infrared
Multi-spectral
Texton
Color

ABSTRACT

Visual saliency detection has been applied in many tasks in the fields of pattern recognition and computer vision, such as image segmentation, object recognition, and image retargeting. However, the accurate detection of saliency remains a challenge. The reasons behind this are that: (1) well-defined mechanism for saliency definition is rarely established; and (2) supporting information for detecting saliency is limited in general. In this paper, a multi-spectrum based saliency detection algorithm is proposed. Instead of only using the conventional RGB information as what existing algorithms do, this work incorporates near-infrared clues into the detection framework. Features of color and texture from both types of image modes are explored simultaneously. When calculating the color contrast, an effective color component analysis method is employed to produce more precise results. With respect to the texture analysis, texton representation is adopted for fast processing. Experiments are done to compare the proposed algorithm with other 11 state-of-the-art algorithms and the results indicate that our algorithm outperforms the others.

1. Introduction

The human vision system pays no equal attention to what they see in the world (James, 1890). To illustrate this point, the top three images in Fig. 1 are employed. Most people would probably say "a tower" or "a tree on the mountain" when looking at these images. Each description undoubtedly covers the most important content in the corresponding image. Actually in this procedure, we unconsciously focus our interest on the salient objects and ignore the other things. This ability of human to "*withdraw from some things in order to deal effectively with others*" (James, 1890) is called attention. The mechanism behind visual attention has been studied intensively in different areas for a long time, but its detection still remains a challenge (Cheng et al., 2011). In this paper, the effort is mainly towards the aspect of computer vision where attention is approached by saliency detection assuming there are one or several salient objects in the examined image.[1]

Visual saliency has found its use in a variety of applications including image segmentation (Ko and Nam, 2006), object recognition (Rutishauser et al., 2004; Walther et al., 2002), image retargeting (Avidan and Shamir, 2007; Rubinstein et al., 2008; Zhang et al., 2009), thumbnailing (Suh et al., 2003) and retrieval (Chen et al., 2009). Existing approaches of saliency detection can be divided mainly into three steps, which are *feature extraction*, *saliency computation* and *saliency map representation*. Features such as color, intensity, orientation and motion are first extracted as basic elements for speculating on saliency. Then a computational method is applied to these features to calculate each pixel's degree-to-be-salient. In the end, a normalization process is conducted to scale the saliency value to the range between 0 and 1. The saliency map can then be displayed as a gray scale image. The whiter a pixel appears, the higher the salient level it has. A typical example of saliency detection results is illustrated in the middle row of Fig. 1.

Approaches for saliency detection can be categorized as model based and computation based. Model based algorithms follow a top-down paradigm. Probably the most influential one of this kind is the selective attention model introduced by Koch and Ullman (1985). They built the saliency map by winner-take-all network based on an early representation mechanism. Itti et al. (1998) were inspired by the theory of visual receptive fields, which states that typical visual neurons are most sensitive in center-surround regions. With this assumption, a set of linear operations at multi-scale were applied to generate visual features producing saliency maps. Walther et al. (2002) extended this method to object recognition and showed an encouraging result which is also biologically plausible. Hou and Zhang (2007) analyzed the log-spectrum of the input image to see whether the distribution obeys the spectral residual model. They assumed that the spectral residual serves as

* Corresponding author. Tel.: +86 29 8888 8837; fax: +86 29 8888 9302.
E-mail address: pingkun@ieee.org (P. Yan).

[1] In this paper, our assumption is to detect the salient objects in an input image assuming there are one or several salient ones in the image. We don't distinguish between the whole object and its parts. This assumption is adopted by most saliency detection papers and almost all ground truths are constructed by this principle.

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2012.06.002

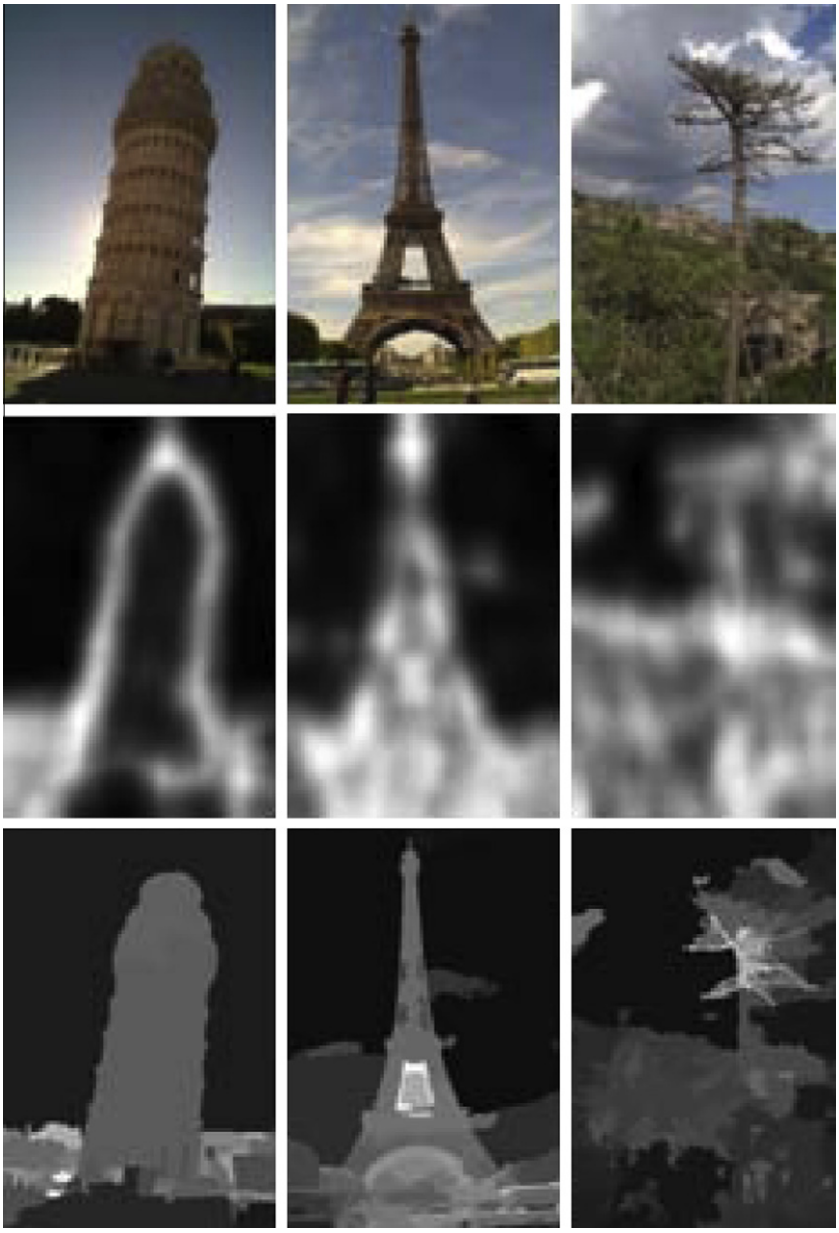

Fig. 1. Saliency detection results. From top to bottom, each row represents the input images, the saliency maps detected by Seo's algorithm (Seo and Milanfar, 2009) and by the proposed algorithm, respectively.

the innovation of the image and can be borrowed to calculate saliency.

As for computational approaches, saliency is defined as contrast of features such as color, motion, orientation, intensity, etc. (Ma and Zhang, 2003; Zhai and Shah, 2006). Liu et al. (2007) proposed to learn a Conditional Random Field to combine features for salient object detection. They also constructed a database containing a large number of images with squared ground truth labels. This database has been widely spread and accepted from then on. Harel et al. (2007) dealt with saliency by graph based method. It consists of two steps: constructing activation maps and then normalizing them by a combination with other maps. Achanta et al. (2009) tackled the problem of saliency detection within frequency domain by applying several difference of Gaussian (DoG) filters of different deviations. These filters are band-pass and can cover the salient region, not just highlighting on edges. Goferman et al. (2010) presented a new definition of saliency, called context-aware saliency which detects regions representing the scene instead of traditionally the dominant object. Achanta et al. (2008) determined the saliency value using the contrast of luminance and color between the inner region and outer region that includes the examined pixel. Kadir and Brady (2001) treat saliency, scale, and image content as interrelated three aspects and tackle them at the same time. Later, this method is modified by Shao et al. (2007) in sampling strategy and calculation of inter- and intra- scale saliency and used in an image retrieval application Shao and Brady (2006). Cheng et al. (2011) also employed the low-level features to define histogram based contrast and region based contrast. To speed up the process, $L^*a^*b^*$ color space is divided into 12^3 bins and a smoothing is followed to reduce noisy artifacts.

Though so many algorithms have been proposed and contributed to this task, the state-of-the-art performance is not satisfying. As illustrated in Fig. 1, we can get a glimpse of the limitation existing in the previous algorithms: the obtained saliency map is

inconsistent. The same object should have an identical salient level in an ideal case. However, this constraint is not satisfied by most the existing algorithms. Some of them generate results with only enhanced boundary maps instead of object regions, while others with fuzzy, low resolution maps do not reflect the image details. For example, the Eiffel Tower in the middle column of Fig. 1 is the salient object and the corresponding area is supposed to have the same saliency level. But the detection results show that the area of Eiffel Tower is not equally salient (with different intensity) by existing method (Seo and Milanfar, 2009). This leads to the inconsistent saliency detection results.

To deal with the problem of inconsistence, in this paper, a novel contrast based saliency detection algorithm is presented. Our contributions are threefold:

(1) Multi-spectral information of RGB and near-infrared bands is incorporated. Existing algorithms for saliency detection are based on the R, G, and B color channels. However, the three bands can provide only limited information for calculating contrast (see Section 3 in the contrast experiments). On the other hand, near-infrared is proved to be differentially informative than the traditional three color channels. It has a weaker dependence on R, G and B than they do to each other, and can provide more clues for recognition and classification (Brown et al., 2011). Therefore, near-infrared band is introduced into saliency detection. To the best of the authors' knowledge, this is the first work to integrate near-infrared with R, G, B bands to fulfill the saliency detection task.

(2) An alternative opponent color analysis method is applied for representing color information. In the early processing of human vision system, opponent colors are present in the form of achromatic (luminance) and opponent (red-green, blue-yellow) parts Ruderman et al., 1998. This can be interpreted as efficient coding – opponent colors decorrelate the signal arriving at the L, M and S photoreceptors (Brown et al., 2011). Inspired by this fact, many researchers employed opponent color (e.g., L*a*b*) to explore the information hidden in images (van de Sande et al., 2010; Ruderman et al., 1998). In this paper, an alternative method for measuring the color contrast is employed to calculate saliency. We make use of the Principle Component Analysis (PCA) to decorrelate the 4-dimentional RGBE (red, green, blue and near-infrared) channels to get a more distinctive and compact descriptor (Brown et al., 2011). It is found out that this new color space transformation with orthogonal basis vectors is more effective than the RGB representations.

(3) Texture features are considered for saliency detection. Humans perceive the world through not only the color information, but also other clues such as texture, depth, shadow and motion. Surprisingly, few works in saliency detection incorporate information other than color. In this work, texture is involved in the saliency detection procedure. By employing texton theory (Malik et al., 2001), texture is effectively expressed and represented. Then they are treated as elemental information for calculating contrast. Experimental results demonstrate that it is more informative than only using the color bands.

The general framework of the proposed algorithm in this paper is as follows. Firstly, a 4-dimentional color vector representation for each pixel is constructed and PCA analysis is applied to get more discriminative description. Secondly, textons are trained over a set of natural images and then each pixel in the input image is described by its nearest textons. After that, saliency maps with respect to color and texture are calculated respectively based on region contrast. To be more specific, saliency maps are computed on color contrast, texton contrast in the regular RGB image and texton contrast in the near-infrared image. Finally, three maps are fused and normalized to obtain the final result. Sample results of the proposed algorithm are shown in the bottom row of Fig. 1.

The remainder of this paper will clarify the proposed algorithm step by step. In Section 2, details of the presented algorithm are described to show how the improvement is achieved. In Section 3, experimental results are shown to validate the proposed method. Discussions and conclusions are made in Section 4.

2. Multi-spectral saliency

Motivated by the observation that the cortical cells respond highly to contrast stimulus in their receptive fields (Reynolds and Desimone, 2003), many computational algorithms are developed to detect saliency based on color contrast. Among those works, most are based on RGB or L*a*b* color channels while textures are rarely considered. This work integrates the multi-spectral color bands and texture to get more informative clues.

2.1. Near-infrared

Humans perceive the world due to the existence of visible light. Its wavelength roughly ranges from 380 nm to 740 nm (Wikipedia, 2012). Besides visible light, there are other electromagnetic waves that cannot be seen by human eyes while they have special physical characteristics different with visible light and can provide abundant information for a variety of tasks. Near-infrared is such a kind with wavelength of 800–2500 nm and increasingly applied in computer vision applications (Brown et al., 2011; Raghavachari, 2001). With a special camera capable of capturing the near-infrared reflection from real world, images can be taken to interpret the different properties of objects under particular circumstances. Motivated by this fact, in this paper the near-infrared information is included for the calculation of saliency in the hope of providing more distinguishable clues.

2.2. Color contrast

Human visual system is only biologically sensitive to visible light. The mixture of the three primary colors R, G and B within this range makes the world so colorful. However, there are other rays such as near-infrared and far-infrared, beyond the visibility of human eyes but can provide more discriminative information for the task of tracking, recognition and classification (Achanta and Susstrunk, 2010; Zhang et al., 2007; Hallaway et al., 2005). Due to the great success of multi-spectral SIFT (Brown et al., 2011), near-infrared information is incorporated in the saliency detection context.

Instead of using the RGBE color channels directly, this work introduces the opponent color decorrelation by PCA, which is proved to be more effective and discriminative (Brown et al., 2011). Suppose the pixels in an image are denoted as $\{\boldsymbol{x}_n\}_{n=1,\ldots,N}$, where N is the number of pixels in the image and $\boldsymbol{x}_n$ is a 4-dimensional color vector $[r, g, b, e]$. The covariance matrix of the pixel data points in the input image is

$$M = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^T, \tag{1}$$

where $\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{n=1}^{N} x_n$. By calculating the eigenvectors of M, we can obtain a decorrelated and reduced expression on four orthogonal basis vectors. This paper chooses the first three projections representing 98% variance of the data. This means for a specific pixel i in the input image, its color vector is defined as $C_i = [\lambda_{i1}, \lambda_{i2}, \lambda_{i3}]$. Its corresponding saliency value is

$$S(C_i, R_i) = \sum_{j \neq i, j \in R_i} d_c(i,j), \tag{2}$$

where R_i is the neighborhood supporting the saliency of pixel i, j is the pixel within R_i and $d_c(i,j)$ is the distance between C_i and C_j. Most of the time, the neighborhood area is selected as a large area surrounding the pixel, sometimes even the whole image. This makes the process computationally intensive. To speed up the process, the reduced color space is first clustered into k_c groups. Each pixel's color is then represented by its group mean. In this way, calculating distances between $256 \times 256 \times 256 \times 256$ colors can be reduced to between k_c color prototypes. Based on this, a $k_c \times k_c$ dictionary D_c is acquired with its element $D_c(i,j)$ being the distance between the ith and jth color prototypes. When computing the distance between two colors, their group centers are first identified. Then we only need to look up the dictionary to find the corresponding distance. According to this color quantization, Eq. (2) can be changed to

$$S(C_i, R_i) = \sum_{j \neq i, j \in R_i} \hbar_{\varphi(j)} D_c(\varphi(i), \varphi(j)), \tag{3}$$

where $\phi(i)$ is the function mapping pixel i to its corresponding group mean. $\hbar_{\phi(j)}$ is the frequency of $\phi(j)$, and $\phi(\cdot) \in \{1, \ldots, k_c\}$.

2.3. Texture contrast

Texture is useful for a variety of visual processing applications. To analyze texture from images, a filter bank with different parameters (including orientation, scale, etc.) is usually applied to the examined pixel and its neighborhood. This work adopts the algorithm proposed by Malik et al. (2001) to incorporate the texture information into our saliency detection procedure. The filter bank used is based on "rotated copies of a Gaussian derivative and its Hilbert transform" (Malik et al., 2001). To be more specific, the two classes of filters are

$$\begin{aligned} f_1(x,y) &= \frac{d^2}{dy^2}\left(\frac{1}{C}\exp\left(\frac{y^2}{\sigma^2}\right)\exp\left(\frac{x^2}{\ell^2\sigma^2}\right)\right), \\ f_2(x,y) &= Hilbert(f_1(x,y)), \end{aligned} \tag{4}$$

where σ is the scale, ℓ is the ratio of the filter, and C is a normalization constant. In this way, each pixel can be characterized by a vector of filter outputs. Similarly, the saliency of pixel i with respect to texture is defined as

$$S(T_i, R_i) = \sum_{j \neq i, j \in R_i} d_t(i,j), \tag{5}$$

where T_i is the texture vector of pixel i containing the outputs of the filter bank. Since the vectorized texture descriptions are numerous, the cost of calculation is expensive. To deal with this problem, the concept of texton is introduced. k_t textons are first trained from a set of natural images. Then the nearest texton is assigned to each pixel. The mapping from pixel to texton provides us with an effective tool to cope with texture in a discrete manner. Then Eq. (5) can be changed to

$$S(T_i, R_i) = \sum_{j \neq i, j \in R_i} \hbar_{\phi(j)} D_t(\phi(i), \phi(j)), \tag{6}$$

where these notations have a similar meaning with those in Section 2.2 but they reflect the texture information.

2.4. Region based enhancement

Calculating the saliency value of each pixel independently would result in a fuzzy map. This is the major drawback of existing saliency algorithms. The reason behind this phenomenon is that the correlation between pixels is not considered properly. In order to get a more consistent saliency map and inspired by Cheng et al. (2011), a region based method is employed, which means that saliency is computed region by region. According to Cheng et al. (2011), humans are prone to pay more attention to regions instead of pixels. This phenomenon is heuristically right by human experience. Besides, region based method is more robust to noise and computationally efficient than pixel based one.

In this work, regions are obtained by a segmentation algorithm. For the obtained region R_k, calculate its saliency value $S(R_k)$ by contrast to all other regions in the image:

$$\begin{aligned} S(R_k) &= \sum_{i \neq k} [\lambda_c d_c(R_k, R_i) + \lambda_{t-rgb} d_{t-rgb}(R_k, R_i) + \lambda_{t-nir} d_{t-nir}(R_k, R_i)] \\ &= \sum_{i \neq k} \sum_{j \in R_k} [\lambda_c S_c(C_j, R_i) + \lambda_{t-rgb} S_{t-rgb}(T_j, R_i) + \lambda_{t-nir} S_{t-nir}(T_j, R_i)], \end{aligned} \tag{7}$$

where $d(R_k, R_i)$ is the contrast between region R_k and R_i with the subscript c, t-rgb, t-nir respectively indicating the color, texture of regular image and texture of near-infrared image. $\lambda_c + \lambda_{t-rgb} + \lambda_{t-nir} = 1$ is satisfied with each indicating their corresponding weight.

According to the Gestalt laws (Koffka, 1995), when looking at an image, there are one or several gravity centers about which the visual form is organized. This suggests that regions surrounding the focus should be explored more than those far away (Goferman et al., 2010). Therefore, a further spatial constraint is imposed on the saliency detection function

$$\begin{aligned} S(R_k) &= \sum_{i \neq k} \exp\{-\rho_s(k,i)\}[\lambda_c d_c(R_k, R_i) + \lambda_{t-rgb} d_{t-rgb}(R_k, R_i) + \lambda_{t-nir} d_{t-nir}(R_k, R_i)] \\ &= \sum_{i \neq k} \sum_{j \in R_k} \exp\{-\rho_s(k,i)\}[\lambda_c S_c(C_j, R_i) + \lambda_{t-rgb} S_{t-rgb}(T_j, R_i) + \lambda_{t-nir} S_{t-nir}(T_j, R_i)], \end{aligned} \tag{8}$$

where $\rho_s(k,i)$ is the spatial distance between the centers of region k and i.

3. Experiments

In order to evaluate the proposed algorithm, 11 image pairs were selected from the publicly available dataset provided by Brown et al. (2011). These images are photographed by several modified SLR cameras with their infrared blocking filters removed. Each pair contains a RGB regular image and a corresponding near-infrared version. In each image, there is one dominant object and this is suitable for our saliency detection task. Then we employ five students major in computer vision to label the salient object in the image. Their labeled results are finally voted by Winner-Takes-All principle to select one ground truth segmentation for each image, examples of which are illustrated in Fig. 2.

In our experiments, the mean shift algorithm (Raghavachari, 2001) is employed to get segmented regions in the beginning. The three weights $\lambda_c, \lambda_{t-rgb}, \lambda_{t-nir}$ are set as 0.7, 0.15, 0.15 empirically. To prove the effectiveness of the proposed algorithm, this paper compares experimental results with those of 11 state-of-the-art algorithms. They are respectively FT (Achanta et al., 2009), LC (Zhai and Shah, 2006), IT (Itti et al., 1998), HC (Cheng et al., 2011), AC (Achanta et al., 2008), CA (Goferman et al., 2010), SeR (Seo and Milanfar, 2009), SR (Hou and Zhang, 2007), SUN (Zhang et al., 2008), and MSS (Achanta and Susstrunk, 2010). The principle for selecting these algorithms is similar to that of Cheng et al. (2011). To be more specific, their recency, variety and popularity are considered together.

All these algorithms are used to compute saliency maps for the 11 images. Our algorithm is implemented in Matlab. The other algorithms are downloaded from the authors' homepage. From Fig. 3, it shows that the saliency maps produced by our algorithm are much clearer than others. This means the salient area can be easily distinguished with the background area. The reason is that the contrast based on color component is more distinguishable than other representations. Therefore, the details of the image can be well reflected and a fuzzy map is avoided.

Fig. 2. Sample images in the experiments. Each pair contains a regular RGB image, a near-infrared image and their ground truth label.

Fig. 3. Visual comparison of saliency maps. (a) original image, saliency maps produced by (b) FT (Achanta et al., 2009), (c) LC (Zhai and Shah, 2006), (d) IT (Itti et al., 1998), (e) HC (Cheng et al., 2011), (f) AC (Achanta et al., 2008), (g) CA (Goferman et al., 2010), (h) SeR (Seo and Milanfar, 2009), (i) SR (Hou and Zhang, 2007), (j) SUN (Zhang et al., 2008), (k) MSS (Achanta and Susstrunk, 2010), and (l) the proposed multi-spectral method.

Besides, our saliency map is more consistent than others. For one identical object in the input image, its salient value is prone to be the same. This is because our saliency value is calculated by region. For pixels in one region, they are with the same saliency value. This principle can restrain noisy saliency values of individual pixels greatly. Fig. 4 shows all the images used in the experiments, their ground truth labels, the computed saliency maps by our algorithm, and the computed saliency maps without region based enhancement. From Fig. 4, it can be observed that the results by the presented algorithm are promising. It is obvious that the obtained results without region based enhancement are fuzzy and inaccurate. But with the region based constraints, the results will be improved a lot.

To evaluate the results in a qualitative way, precision and recall (P–R) curves, which are a standard technique for information retrieval community, are adopted. According to Bowyer et al. (2001), Abdou and Pratt (1979), and Martin et al. (2004), precision and recall are defined as

$$precision = \frac{TP}{TP + FP}, recall = \frac{TP}{TP + FN}, \tag{9}$$

where TP is true positive, FP false positive and FN false negative. Saliency maps are thresholded by different values to calculate these indexes and the P–R curves are then plotted. From those curves in Fig. 5, it concludes that the proposed algorithm is superior to others. Achieving at the same precision value, the proposed algorithm can detect more salient regions; with the same recall value, the proposed algorithm is more accurate.

To further explore the effectiveness of incorporating the near-infrared and texture clues, two sets of experiments were carried out. The first set is to test the performance with and without near-infrared color band. In this case, each image is processed twice based on the RGB and RGBE color bands respectively. Then the averaged precision and recall values are calculated. In order to capture the tradeoff between the two indexes, F-measure (Van Rijsbergen, 1979), which is defined as

$$F - measure = \frac{precision \times recall}{(1 - \alpha) \times precision + \alpha \times recall}, \tag{10}$$

is employed to evaluate the performance in a compromised manner. Here α is set as 0.5 according to Martin et al. (2004). Fig. 6(a) illustrates the results. It shows that with the addition of near-infrared color band, the averaged precision and F-measure are higher than without it.

The second test evaluates the influence of texture on the saliency detection results. Experiments are done in condition of texture and without texture information. Fig. 6(b) shows the results. It is obvious that texture information can improve the performance of saliency detection.

4. Discussion and conclusions

In this paper, a multi-spectral based saliency detection algorithm is proposed. The algorithm incorporates the near-infrared and the texture clues into the procedure of saliency detection. As for color contrast, PCA analysis is employed to get an effective representation. This representation makes the color vector more descriptive and distinctive. In terms of texture, the widely accepted texton theory is adopted to discrete the continuous texture space. By this means, the presented algorithm can provide more informative clues for saliency detection. Besides, a region based enhancement is also incorporated into the saliency calculation process. We believe that low level features of color and texture alone cannot interpret the superior performance of the proposed algorithm. Coupled with the location constraint and region based enhancement, the resulting performance appears to be so promising. The two factors are correlated and co-influenced. We then compare

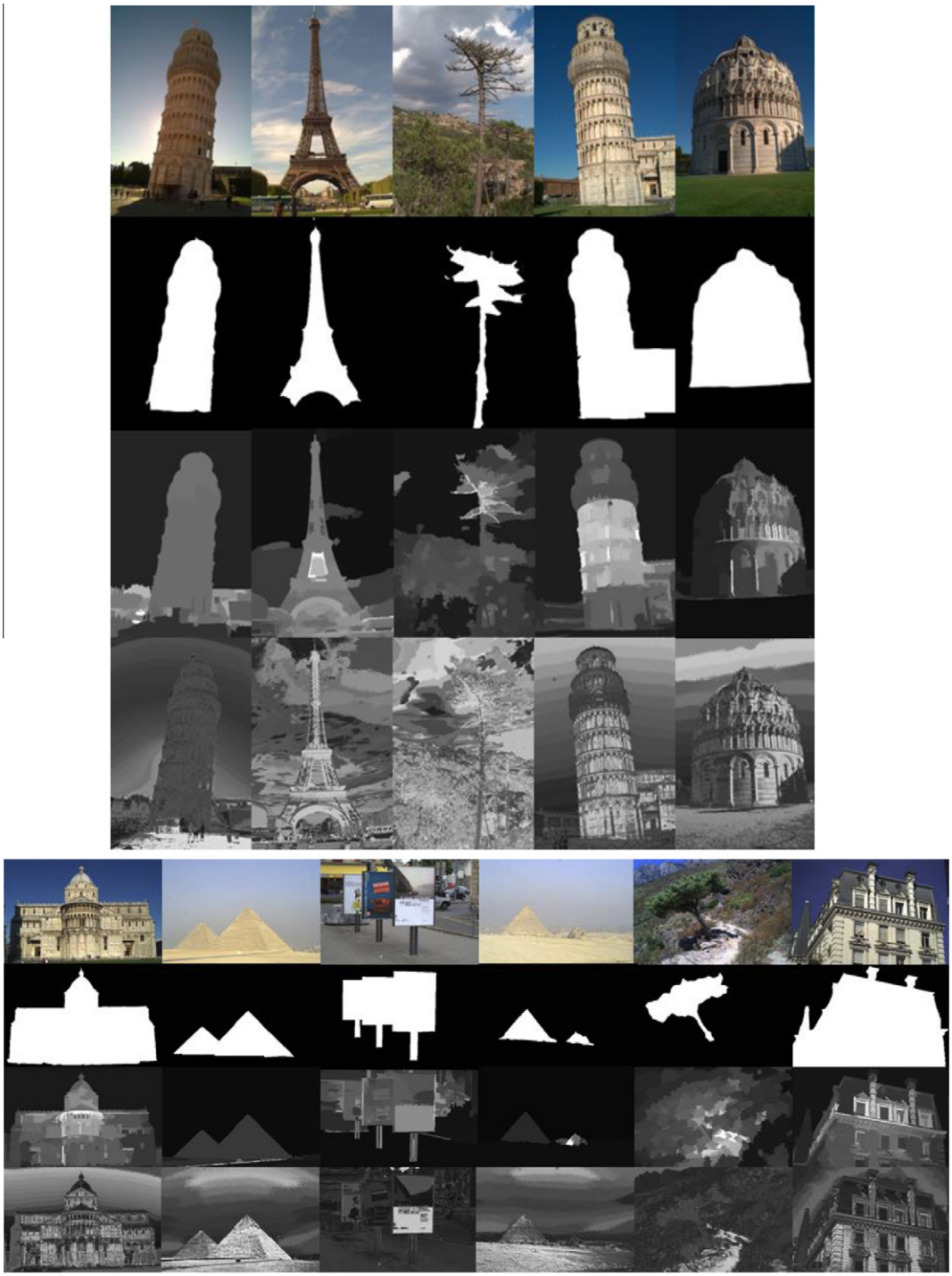

Fig. 4. First row: the images used in the experiments. Second row: ground truth labels. Third row: the saliency detection results by our algorithm. Fourth row: the saliency detection results without region based enhancement.

the proposed algorithm with other 11 state-of-the-art ones and experimental results demonstrate that our algorithm performs better.

The following will discuss several issues in the algorithm formulation and experimental process. The first one is the number of color and texton groups. In the experiments, we have tried to divide the color and texture space into 32, 64 and 128 groups. It is found that 32 are too small to be distinctive and 128 are computationally expensive. Therefore, 64 is a compromise with moderate computational cost and adequate distinctiveness.

The second one is the image number in experiments. There are 400+ images in Brown's dataset (Brown et al., 2011), which are photographed by several modified SLR cameras with the infrared blocking filter removed, but only 11 are chosen for our experiments. This is not a large number because most of them are not suitable for the saliency detection task. The contents of these images, shown in Fig. 7, have no obvious visual focus. However, our algorithm outperforms the others not a bit from the subjective evaluation (see Fig. 3) and the precision-recall curves (see Fig. 5). We don't think it is a coincidence. In the next step, we plan to build a dataset containing

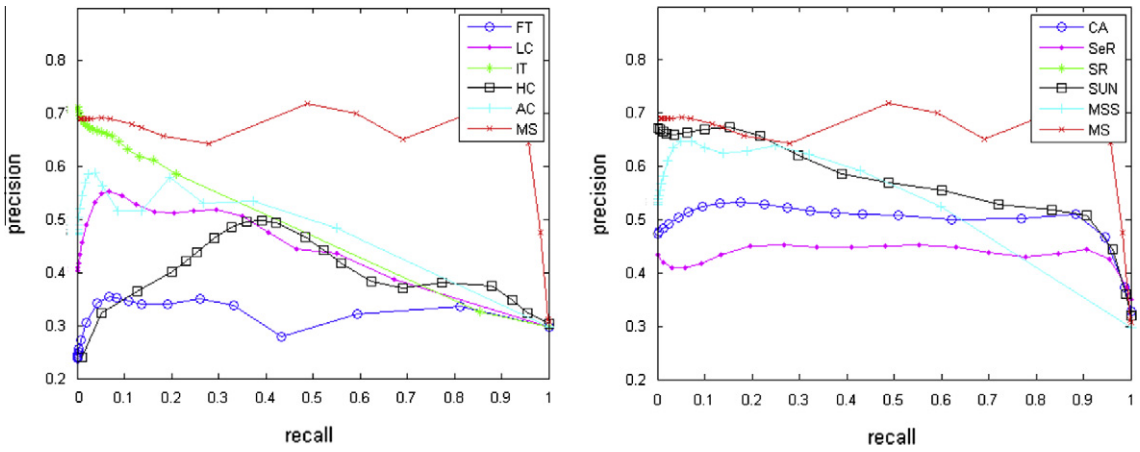

Fig. 5. Precision-recall curves for algorithms in the experiments. They are respectively FT (Achanta et al., 2009), LC (Zhai and Shah, 2006), IT (Itti et al., 1998), HC (Cheng et al., 2011), AC (Achanta et al., 2008), CA (Goferman et al., 2010), SeR (Seo and Milanfar, 2009), SR (Hou and Zhang, 2007), SUN (Zhang et al., 2008), MSS (Achanta and Susstrunk, 2010), and the proposed multi-spectral MS algorithm.

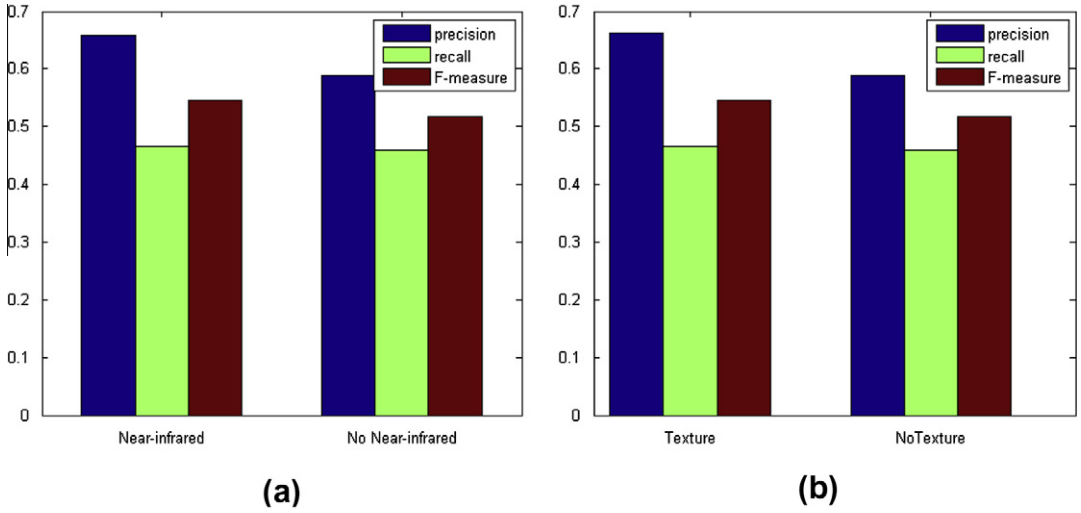

Fig. 6. Saliency detection performance (a) with and without near-infrared color band; (b) with and without texture clue.

Fig. 7. Image samples from dataset (Brown et al., 2011).

500+ image pairs with ground truth labels to evaluate our algorithm. Furthermore, we want to make the dataset a benchmark for saliency detection and segmentation. This work has been started.

Acknowledgements

This Project is supported by the National Basic Research Program of China (973 Program) (Grant No. 2011CB707104), the National Natural Science Foundation of China (Grant Nos. 61105012, 61172142, 61172143, 61125106, and 91120302), the Natural Science Foundation Research Project of Shaanxi Province (Grant No. 2012JM8024), the 50th China Postdoctoral Science Foundation (Grant No. 2011M501487) and the Special Fund for Basic Scientific Research of Central Colleges, Chang'an University (Grant No. CHD2011JC124).

References

Abdou, I., Pratt, W., 1979. Quantitative design and evaluation of enhancement/thresholding edge detectors. Proc. IEEE 67 (5), 753–763.
Achanta, R. Susstrunk, S., 2010. Saliency detection using maximum symmetric surround. In: Proc. IEEE Internat. Conf. Image Process., pp. 2653–2656.
Achanta, R., Estrada, F., Wils, P., Süsstrunk, S., 2008. Salient region detection and segmentation. Comput. Vision Systems 5008, 66–75.

Achanta, R., Hemami, S., Estrada, F. Susstrunk, S., 2009. Frequency-tuned salient region detection. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1597–1604.
Avidan, S., Shamir, A., 2007. Seam carving for content-aware image resizing. ACM Trans. Graphics 26 (2).
Bowyer, K., Kranenburg, C., Dougherty, S., 2001. Edge detector evaluation using empirical roc curves. Comput. Vision Image Underst. 84 (1), 77–103.
Brown, M., Susstrunk, S., 2011. Multi-spectral SIFT for scene category recognition. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition.
Chen, T., Cheng, M.-M., Tan, P., Shamir, A., Hu, S.-M., 2009. Sketch2Photo: Internet image montage. ACM Trans. Graphics 28 (5).
Cheng, M.-M., Zhang, G.-X., Mitra, N.J., Huang, X., Hu, S.M., 2011. Global contrast based salient region detection. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 409–416.
Goferman, S., Zelnik-Manor, L., Tal, A., 2010. Context-aware saliency detection. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2376–2383.
Hallaway, D., Hollerer, T., Feiner, S., 2005. Coarse, inexpensive, infrared tracking for wearable computing. In: Proc. IEEE Wearable Computers, pp. 69–78.
Harel, J., Koch, C., Perona, P., 2007. Graph-based visual saliency. Adv. Neural Inform. Process. Syst., 545–552.
Hou, X., Zhang, L, 2007. Saliency Detection: A spectral residual approach. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–8.
Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Machine Intell. 20 (11), 1254–1259.
James, W., 1890. The Principles of Psychology. Holt, New York.
Kadir, T., Brady, M., 2001. Scale, saliency and image description. Internat. J. Comput. Vision 45 (2), 83–105.
Ko, B.C., Nam, J.-Y., 2006. Object-of-interest image segmentation based on human attention and semantic region clustering. J. Optical Soc. Amer. A 23 (10), 2462–2470.
Koch, C., Ullman, S., 1985. Shifts in selective visual attention: Towards the underlying neural circuitry. Human Neurobiology 4 (4), 219–227.
Koffka, K., 1995. Principles of Gestalt Psychology. Routledge & Kegan Paul.
Liu, T., Sun, J., Zheng, N.-N., Tang, X, Shum, H. -Y. 2007. Learning to detect a salient object. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–8.
Ma, Y.-F., Zhang, H.-J., 2003. Contrast-based image attention analysis by using fuzzy growing. In: Proc. ACM Multimedia, pp. 374–381.
Malik, J., Belongie, S., Leung, T., Shi, J., 2001. Contour and texture analysis for image segmentation. Internat. J. Comput. Vision 43 (1), 7–27.
Martin, D.R., Fowlkes, C.C., Malik, J., 2004. Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Trans. Pattern Anal. Machine Intell. 26 (5), 530–549.
Raghavachari, R., 2001. Near-Infrared Applications in Biotechnology. Marcel-Dekker, New York, NY.
Reynolds, J., Desimone, R., 2003. Interacting roles of attention and visual salience in v4. Neuron 37 (5), 853–863.
Rubinstein, M., Shamir, A., Avidan, S., 2008. Improved seam carving for video retargeting. ACM Trans. Graphics 27 (3).
Ruderman, D., Cronin, T., Chiao, C., 1998. Statistics of cone responses to natural images: Implications for visual coding. J. Optical Soc. Amer. 15 (8), 2036–2045.
Rutishauser, U., Walther, D., Koch, C., Perona, P., 2004. Is bottom-up attention useful for object recognition? In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol. II, pp. 37–44.
Seo, H.J., Milanfar, P., 2009. Static and space-time visual saliency detection by self-resemblance. J. Vision 9 (12), 1–27.
Shao, L., Brady, M., 2006. Specific object retrieval based on salient regions. Pattern Recognition 39 (10), 1932–1948.
Shao, L., Kadir, T., Brady, M., 2007. Geometric and photometric invariant distinctive regions detection. Inf. Sci. 177 (4), 1088–1122.
Suh, B., Ling, H., Bederson, B.B., Jacobs, D.W., 2003. Automatic thumbnail cropping and its effectiveness. In: Proc. 16th Annual ACM Symposium on User Interface Software and Technology, pp. 95–104.
van de Sande, K., Gevers, T., Snoek, C., 2010. Evaluating color descriptors for object and scene recognition. IEEE Trans. Pattern Anal. Machine Intell. 32 (9), 1582–1596.
Van Rijsbergen, C., 1979. Information Retrieval, second ed. Dept. of Computer Science, Univ. of Glasgow.
Walther, D., Itti, L., Riesenhuber, M., Poggio, T., Koch, C., 2002. Attentional selection for object recognition – A gentle way. Biologically Motivated Comput. Vision 2525, 251–267.
<http://en.wikipedia.org/wiki/Near_infrared>.
Zhai, Y., Shah, M., 2006. Visual attention detection in video sequences using spatiotemporal cues. In: Proc. ACM Multimedia, pp. 815–824.
Zhang, L., Wu, B., Nevatia, R., 2007. Pedestrian detection in infrared images based on local shape features. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition.
Zhang, L., Tong, M.H., Marks, T.K., Shan, H., Cottrell, G.W., 2008. SUN: A Bayesian framework for saliency using natural statistics. J. Vision 8 (7), 1–20.
Zhang, G.-X., Cheng, M.-M., Hu, S.-M., Martin, R.R., 2009. A Shape-preserving approach to image resizing. Comput. Graphics Forum 28 (7), 1897–1906.