Computer Vision and Image Understanding 117 (2013) 1748–1754

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Multi-spectral dataset and its application in saliency detection

Qi Wang, Guokang Zhu, Yuan Yuan *

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

ARTICLE INFO

Article history:
Received 14 July 2012
Accepted 3 July 2013
Available online 26 July 2013

Keywords:
Multi-spectral
Near-infrared
Saliency
Regression model

ABSTRACT

Saliency detection has been researched a lot in recent years. Traditional methods are mostly conducted and evaluated on conventional RGB images. Few work has considered the incorporation of multi-spectral clues. Considering the success of including near-infrared spectrum in applications such as face recognition and scene categorization, this paper presents a multi-spectral dataset and applies it in saliency detection. Experiments demonstrate that the incorporation of near-infrared band is effective in the saliency detection procedure. We also test the combinational models for integrating visible and near-infrared bands. Results show that there is no single model to effect on every saliency detection method. Models should be selected according to the specific employed method.

1. Introduction

Saliency detection has been a promising topic recently [1–4]. The goal of saliency detection is to extract salient areas from an input image and present the result as a gray scale image. The whiter the pixel seems, the more possible it might be salient. Since the detected saliency map can be utilized in various applications, such as recognition [5], segmentation [6], and tracking [7], research towards this subject has attracted much attention [8–10].

Generally, methods for saliency detection can be categorized into local based and global based schemes [11]. Local based methods calculate a region's saliency according to the contrast to a small neighborhood [12–14]. Global based methods evaluate saliency with respect to the whole image's statistical characteristic [15,16]. Whatever the case is, saliency detection is mostly conducted on natural images taken by ordinary cameras. These cameras can respond to wavelengths from about 390 to 700 nm, which is called the visible spectrum [17]. The obtained images are regular RGB images. As for the electromagnetic spectrums beyond this scope, their information is lost during the imaging process. However, the lost spectrums might be also valuable for vision tasks because the more supporting information we have, the more rationale decisions will be made. This judgment is not only the common sense for humans, but also proved by other applications in computer vision field. For example, after the proposition of SIFT descriptor [18] on gray scale images, CSIFT [19,20] was developed to incorporate the color bands into the descriptor. Then not long ago, MSIFT [21] was presented to include the near-infrared band for a richer descriptor. As for the face recognition research, early work primarily focus on the gray or RGB images. Later, other light bands besides the visible spectrum [22] are involved to eliminate the lighting problem. The same is true for boundary detection [23] and tracking [24] that incorporating more clues will improve the performance. In remote sensing, the spectrum is not limited to one or several bands, but up to a level of tens and hundreds [25–27].

Considering the success of including other light bands besides the visible light in many applications, we construct a multi-spectral dataset containing both near-infrared (NIR) and regular RGB images in this work. Several dataset containing NIR images have been presented before. For example, the PolyU-NIRFD dataset [22] for face recognition, the NIR–RGB dataset [21] for scene categorization. But these datasets are designed for specific purpose. They cannot be readily utilized for saliency detection. To this aim, the presented dataset is constructed in the hope of providing a new platform for saliency research.

The rest of this paper is organized as follows. Section 2 presents the proposed multi-spectral dataset. Section 3 introduces the distinguishable properties of near-infrared band. Section 4 applies the presented dataset in saliency detection. Finally, conclusion is made in Section 5.

2. Multi-spectral dataset

Since more clues are prone to provide richer information, we hope that a camera can capture the NIR and RGB spectrums simultaneously. However, most existing datasets contain images captured from only RGB bands. We cannot get the information of

* Corresponding author.
E-mail addresses: crabwq@opt.ac.cn (Q. Wang), zhuguokang@opt.ac.cn (G. Zhu), yuany@opt.ac.cn (Y. Yuan).

1077-3142/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cviu.2013.07.002

the four bands at the same time. Though the NIR–RGB dataset [21] has images of both bands, each pair of them are taken consecutively with two cameras. This makes the contents of image pairs not the same. When these images are employed, they have to be accurately registered. But the obtained results are still not satisfying because some objects exist in one image but not in the other. Considering this problem, we employ a multi-spectral camera to simultaneously capture the images of the four bands.

The camera we employed is a prism based 2-CCD progressive area scan one, the configuration of which is shown in Fig. 1(a). We can see clearly that the prisms in the camera spit the input light into two channels. One is 400–700 nm visible spectrums of red, green and blue, and the other is 700–1000 nm NIR spectrums. This separation is accurately ensured by the dichroic coatings of the prisms. The splitted spectrums are then responded by two distinct CCDs, each of which is sensitive to a range of wavelengths. Their response curves are shown in Fig. 1(b) and (c). The advantage of this camera is that it can capture two images at the same time and the obtained image pair are with the same scope and content.

With this camera, we took 40 pairs of 512×384 images of indoor and outdoor scenes and each image contains one or several salient objects within it. Then the salient objects in each pair were labeled by 5 graduate students major in computer vision. In this procedure, few instructions were given to the participants except segmenting the salient ones they thought as. This ensures the minimum amount of influence on the participants' labelings due to the unnecessary instructions. Since every individual's perception is different, their labelings differ with each other. To get an unbiased ground truth, we select the common areas of each participant's labeling as the final results. Typical examples are shown in Fig. 2.

3. NIR spectrum

The NIR spectrum is between the visible light band and the thermal infrared band. It has the properties of both visible light and thermal infrared light, but is different from any of them. Firstly, unlike thermal infrared, objects can reflect the NIR light the same way as they do to visible light. Secondly, it is invisible to human eyes like the thermal infrared and reflects an "unseen" characteristic different from visible light.

In order to know the relationship and difference between the RGB and NIR spectrums, we plot their pairwise cooccurrence distributions on the 40 image pairs in the 2D plane. All the RGB and NIR values are normalized to [0,1] and different occurrence frequencies are denoted by different colors. From Fig. 3, it is obvious that the distributions of RG, RB and GB are different from those of RN, GN and BN. The latter ones spread more widely in the 2D plane. This implies that the original visible light of red, green and blue are much higher correlated in a pairwise manner than the NIR with them. The NIR spectrum can provide more different information than the visible spectrum.

To justify this point, we calculate the joint entropy [21] of each two bands as

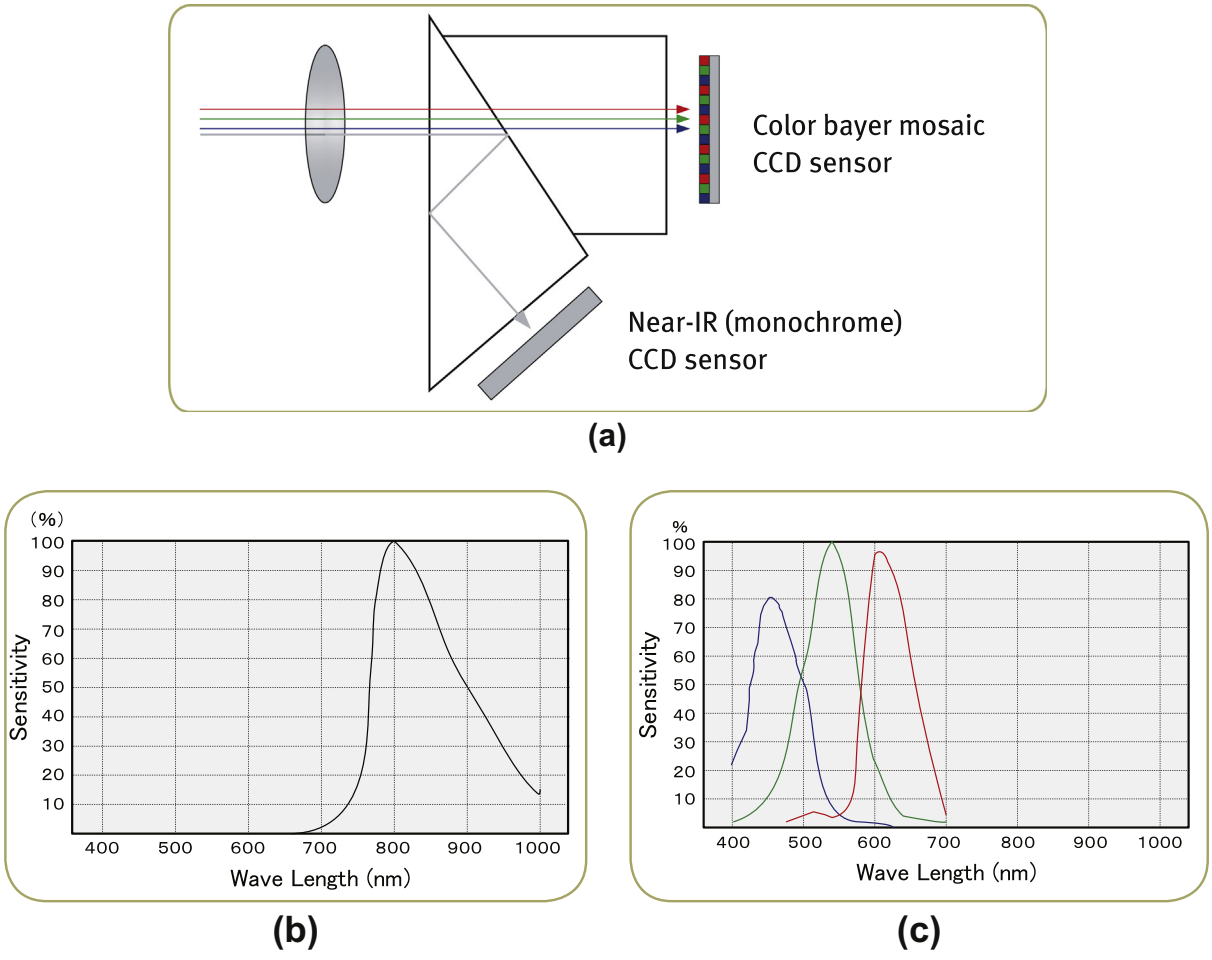

Fig. 1. (a) The mechanism of the camera. (b) The CCD response curve to the NIR spectrum. (c) The CCD response curve to the visible light spectrum. These two figures are cited from [28].

Fig. 2. Example images in the presented dataset. First row: RGB images; Second row: NIR images; Third row: ground truth labelings of salient objects.

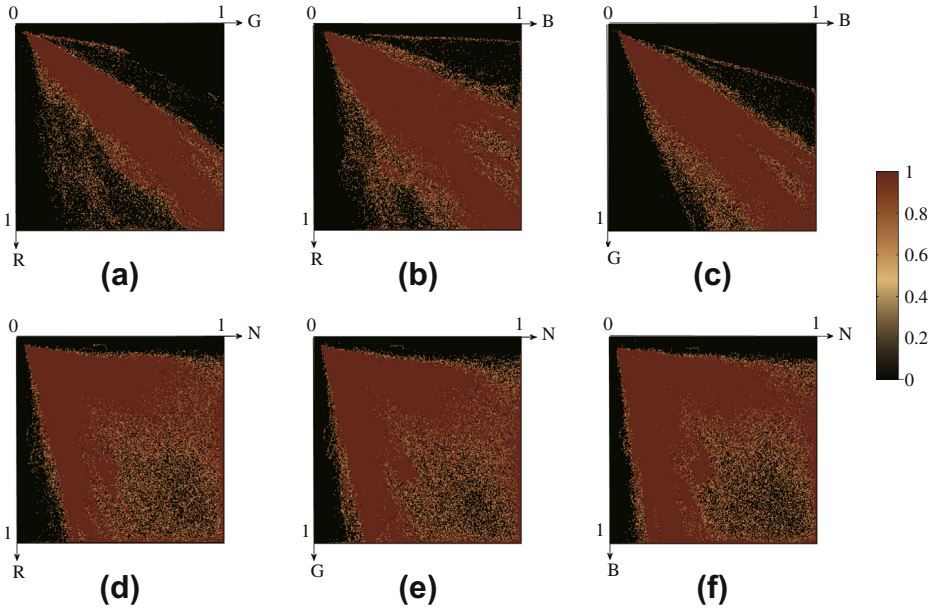

Fig. 3. The cooccurrence distribution of the pairwise bands. This statistic is obtained from the 40 image pairs in the presented dataset.

$$H(X,Y) = \sum_{i,j} -p(x_i,y_j)\log_2 p(x_i,y_j), \tag{1}$$

where X and Y are the examined spectrums, x_i and y_j are the pixel values of the corresponding spectrum image, and $p(x_i,y_j)$ is the probability density. According to the information theory, the entropy is a measure of unpredictability and reflects the information content. The higher $H(X,Y)$ is, the more information is contained in a message. The calculated joint entropy is shown in Table 1. From the table, we can see that the joint entropies with the NIR band are generally higher than those without it. This result demonstrates that the NIR band can provide much abundant information. Trying to utilize it in applications is reasonable.

Table 1
Joint entropy of pairwise bands.

Pairwise bands	RG	RB	GB	RN	RN	BN
Joint entropy	13.488	13.967	13.688	14.361	14.435	14.456

4. Saliency detection

To demonstrate the effectiveness of the presented dataset, we conduct experiments in the application of saliency detection. Saliency maps are firstly extracted from RGB images and NIR images. Then the obtained maps are combined together to get the final results. The purpose of these experiments is to answer the two following questions: 1) whether or not the incorporation of NIR band can improve the saliency detection performance; 2) which kind of models is the best to combine the saliency maps from the two channels.

To answer the first question, we compare the results generated with only RGB band and the results with both RGB and NIR bands. Algorithms employed in this process are all canonical ones in

saliency detection fields. They are AC [29], CA [13], FT [16], HC [11], IT [12], LC [30], MSS [31], RC [11], SR [32] and SUN [33]. To answer the second question, it is more difficult because a traversal test of comparative models is impossible. Considering the initial success of [34], we concentrate on the regression models in this work.

4.1. Evaluation measure

To compare experimental results, an evaluation measure should be firstly specified. In saliency detection field, three metrics are usually employed: *precision*, *recall*, and *F-measure*. They are defined as follows

$$\begin{aligned} &precision = \frac{TP}{TP+FP}, \quad recall = \frac{TP}{TP+FN}, \\ &F\text{-}measure = \frac{precision \times recall}{(1-\alpha)\times precision + \alpha \times recall}, \end{aligned} \tag{2}$$

where *TP* is true positive, *FP* is false positive, and *FN* is false negative. These three metrics are usually utilized in information retrieval community and each of them reflects a different aspect. Precision represents the accuracy, recall represents the detectability, and F-measure is a balance between them. When precision and recall contradict with each other, F-measure is usually employed to represent a compromised measurement [35].

4.2. Regression models

In our processing, the aim is to infer each pixel's saliency value according to its obtained saliency maps from the RGB and NIR bands. Suppose the RGB and NIR saliency are denoted as X_{rgb} and X_{nir}, respectively. The question is how to determine the mapping function f: $(X_{rgb}, X_{nir}) \rightarrow Y$, where Y is the desired saliency value. Three commonly used regression models are employed here. They are *linear regression*, *polynomial regression* and *logistic regression*.

4.2.1. Linear regression.

In linear regression [36], the output variable is a linear combination of input variables. To be specific, the model can be expressed as

$$Y = \alpha_0 + \alpha_1 X_{rgb} + \alpha_2 X_{nir}. \tag{3}$$

The task is to estimate $\{\alpha_i\}_{i=0,1,2}$ from the observed N training points $\{X^n_{rgb}, X^n_{nir}, Y^n\}_{n=1,\ldots,N}$. A special case of linear regression is that the constant coefficient α_0 equals 0. In this case, the output is only the proportional combination of X_{rgb} and X_{nir}, with no translation. The two models are abbreviated as *LinearR-I* and *LinearR-II* in later discussion.

4.2.2. Polynomial regression

In polynomial regression, the independent variables include not only linear terms, but also quadratic and interactive terms. The model is expressed as

$$Y = \alpha_0 + \alpha_1 X_{rgb} + \alpha_2 X_{nir} + \alpha_3 X_{rgb} X_{nir} + \alpha_4 X^2_{rgb} + \alpha_5 X^2_{nir}. \tag{4}$$

The processing is the same as linear regression. According to the known input and output pairs, get an estimation of $\{\alpha_i\}_{i=0,\ldots,5}$ that best fit the training data. This model is denoted as *PolyR* to facilitate the representation.

4.2.3. Logistic regression

For logistic regression [37], the model is defined as

$$f(X) = \frac{e^X}{e^X + 1} = \frac{1}{1 + e^{-X}}, \tag{5}$$

where X represents some set of independent variables, which in this work is defined as

$$X = \alpha_0 + \alpha_1 X_{rgb} + \alpha_2 X_{nir}. \tag{6}$$

The model reflects the nonlinear relationship between the input and output variables, especially emphasizing an approximately linear mapping in the mid-range of input variables and stretching out the extremes exponentially. It is denoted as *LogisticR*.

4.3. Experiments

In this section, intensive comparative experiments are conducted to answer the previously posed two questions. 10 images are selected for training the parameters of the regression models and the other 30 for testing. Since each image contains 512×318 pixels, involving them all from the 10 training images will lead to a great number of training samples. Besides, there are many repeated triplets $\{X^n_{rgb}, X^n_{nir}, Y^n\}$ with the same values if all the pixels are employed, which will lead to a biased model. Therefore, we first resize the training images to 1/4 times of the original size. Then all the pixels ($128 \times 80 \times 10 = 102{,}400$) are utilized as training samples. After all the parameters are learned from the training stage, the remaining 30 images are processed by the acquired models. Typical examples of the regression results are shown in Fig. 4. To quantitatively evaluate the performance of each regression model, we plot the averaged precision, recall and F-measure bars of the original methods and their improvements (in percentage) of the regression models in Fig. 5. Detailed analysis is presented below.

4.3.1. Whether or not the incorporation of NIR band can improve the saliency detection performance?

From Fig. 5, we can see that the incorporation of NIR band makes the three metrics change greatly. Sometimes, they change towards the desired direction. Other times, the inclusion of NIR band makes the indexes drop unsatisfyingly. There are still cases that the precision and recall change oppositely or their improved degrees are different. This leads to a difficult comparison of the regression models. According to [35], F-measure should be referred to in this case because it is a compromise of precision and recall. From Fig. 5 it is evident that for each saliency detection method, there is at least one regression model that can improve the F-measure of the detected results. This means the added NIR band is effective when employing the appropriate model. The maximum improvement is approximately 40% and the minimum is about 5% in our experiments.

4.3.2. Which kind of models is the best to combine the saliency maps from the two channels?

For each saliency detection method, its corresponding regression model performs differently. Some models improve the results, while others may make the results worse. But every method has at least one regression model that can make the original results better. The best regression model for each method is labeled by an purple rectangle. We can see clearly that not every method has an identical best regression model. For CA, FT, HC, LC, RC and SUN, their best performance are achieved on LinearR-II model. For AC and SR, LinearR-I fits them more properly. And for IT and MSS, PolyR is the most appropriate one.

Since each saliency detection method generates a different result, which represents a distinguishable data distribution and paradigm, their best combinational manner might differ with each other. Six methods achieve their excellence on LinearR-II model. This means the LinearR-II is a more generalized and suitable model for saliency detection. No methods perform best on LogisticR model. This means that the model is not fit for these saliency detection

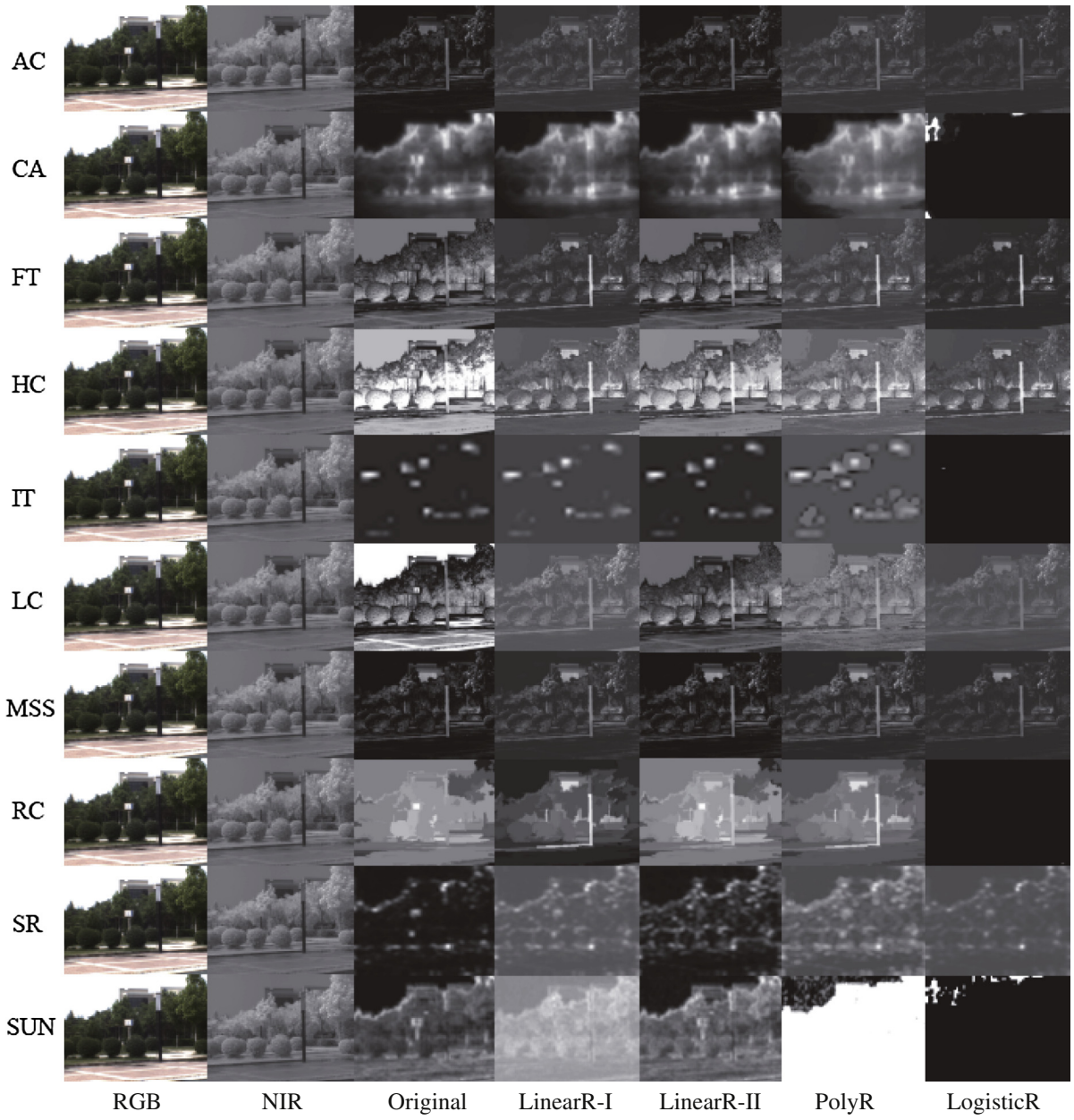

Fig. 4. Typical examples of experimental results. The first column to the seventh column respectively represent the RGB image, NIR image, the results using only the original RGB band, the results of combing RGB and NIR bands by LinearR-I, LinearR-II, PolyR and LogisticR models.

techniques (e.g. several results are all black in the last column of Fig. 4).

4.4. Discussion

In the above Section, experimental results are presented and analyzed. There are still several issues remaining to be discussed in this part.

The first one is about model selection. In this work, we only focus on the conventional regression models to describe the combinational relationship of RGB and NIR bands. However, since there are still other models for the combination of the two bands and we have not tested them all, we do not claim the regression models are the most perfect ones. On the contrary, we think models should be selected according to the specific saliency detection method. This is reasonable to understand because models are built on the actual data. Different methods generate dissimilar saliency maps, which represent distinct data paradigms. Consequently, there is no reason to assume an universal model to all methods.

The second one is about the performance of the employed saliency detection methods. From Fig. 4 and Fig. 5, we can see that the employed canonical methods do not perform well enough as reported in [11]. This is because the presented dataset is a little more complex than the dataset utilized in [11]. The backgrounds usually contain several disturbing objects of different colors. The salient objects are not with the distinguishable fresh color, either. This phenomenon also reveals that the current state-of-the-art methods heavily rely on the color contrast and its robustness deserves to be improved.

The third one is the size of the presented dataset. 40 images are not a large number. This is because the platform for capturing images is not convenient to move around. We only set the environ-

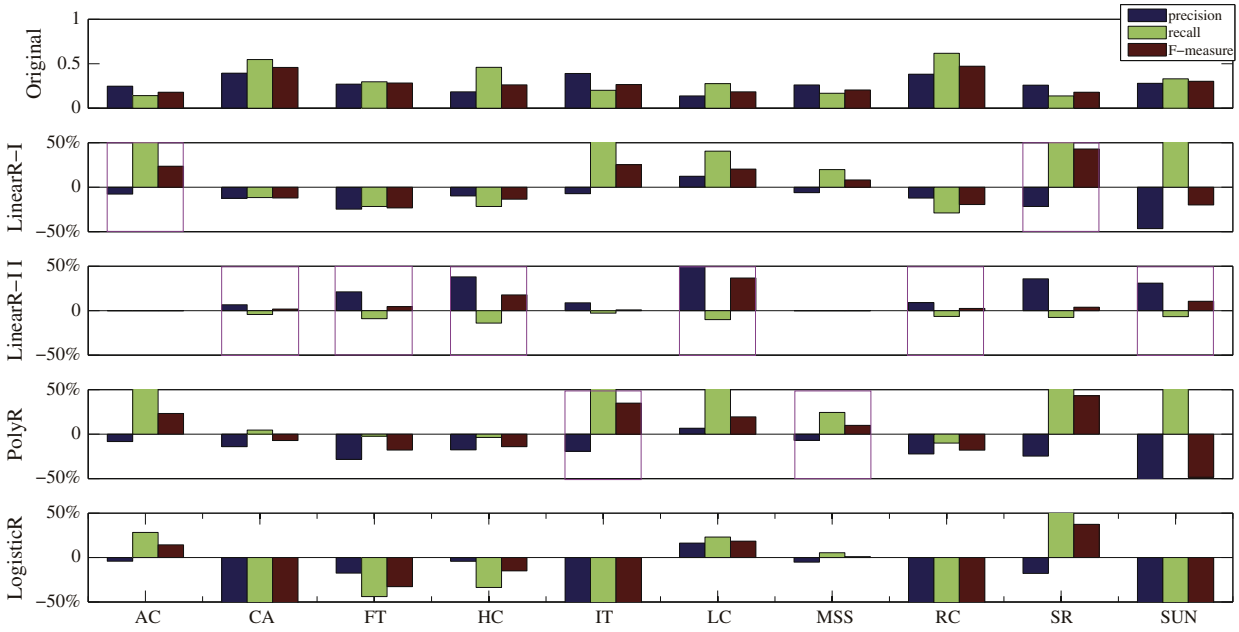

Fig. 5. Quantitative evaluation of different regression models by precision, recall and F-measure. Their results are represented by the improved percentages compared with the original methods using only RGB band. The first row illustrates the original saliency detection results using common RGB images. The second to the fifth rows are the results by combing RGB and NIR bands using LinearR-I, LinearR-II, PolyR and LogisiticR regression models. The results show that for each saliency detection method, (1) there is at least one regression model that can improve the F-measure of the detected results; and (2) its best combinational manner of RGB and NIR bands might differ with other methods.

ment within and not far from the laboratory. Now we are working on transplanting the desktop capturing system to a laptop and taking more images to form a larger dataset. On the other hand, the average F-measure can be improved with a maximum of 40%, which is not a small improvement. We think it is not a coincidence and presents a promising result.

5. Conclusion

In this work, a multi-spectral dataset is presented to serve as a new platform for saliency research. Different from existing ones, our dataset contains pairs of RGB and NIR images. This can provide more valuable information for detecting the salient areas in an image. Experiments demonstrate the effectiveness of the incorporation of NIR band in saliency detection. We also test several regression models for combining the RGB and NIR bands. Results show that it is not appropriate to employ one single model as prototype. The best model should be selected according to the specific method. Future work plans to transplant the image capturing system to a laptop and take more images into the dataset.

Acknowledgment

This work is supported by the State Key Program of National Natural Science of China (Grant No. 61232010), the National Natural Science Foundation of China (Grant No. 61172143 and 61105012), and the Natural Science Foundation Research Project of Shaanxi Province (Grant No. 2012JM8024).

References

[1] T. Jost, N. Ouerhani, R. von Wartburg, R. Müri, H. Hügli, Assessing the contribution of color in visual attention, Comput. Vis. Image Understand. 100 (2005) 107–123.
[2] D. Walther, U. Rutishauser, C. Koch, P. Perona, Selective visual attention enables learning and recognition of multiple objects in cluttered scenes, Comput. Vis. Image Understand. 100 (2005) 41–63.
[3] I. Bogdanova, A. Bur, H. Hügli, P.-A. Farine, Dynamic visual attention on the sphere, Comput. Vis. Image Understand. 114 (2010) 100–110.
[4] Q. Wang, Y. Yuan, P. Yan, X. Li, Saliency detection by multiple-instance learning, IEEE Trans. Cybernetics 43 (2013) 660–672.
[5] D. Walther, L. Itti, M. Riesenhuber, T. Poggio, C. Koch, Attentional selection for object recognition – a gentle way, Biol. Motivated Comput. Vis. 2525 (2002) 251–267.
[6] J. Han, K. Ngan, M. Li, H. Zhang, Unsupervised extraction of visual attention objects in color images, IEEE Trans. Circ. Syst. Video Technol. 16 (2006) 141–145.
[7] V. Mahadevan, N. Vasconcelos, Saliency-based discriminant tracking, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2009, pp. 1007–1013.
[8] R. Perko, A. Leonardis, A framework for visual-context-aware object detection in still images, Comput. Vis. Image Understand. 114 (2010) 700–711.
[9] Y. Sun, R.B. Fisher, F. Wang, H.M. Gomes, A computer vision model for visual-object-based attention and eye movements, Comput. Vis. Image Understand. 112 (2008) 126–142.
[10] T.E. de Campos, G. Csurka, F. Perronnin, Images as sets of locally weighted features, Comput. Vis. Image Understand. 116 (2012) 68–85.
[11] M.M. Cheng, G.X. Zhang, N.J. Mitra, X. Huang, S.M. Hu, Global contrast based salient region detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2011, pp. 409–416.
[12] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 1254–1259.
[13] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 2376–2383.
[14] T. Liu, J. Sun, N.N. Zheng, X. Tang, H.Y. Shum, Learning to detect a salient object, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[15] Y. Zhai, M. Shah, Visual attention detection in video sequences using spatiotemporal cues, in: ACM Multimedia, 2006, p. 815–824.
[16] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2009, pp. 1597–1604.
[17] H.W. Siesler, Y. Ozaki, S. Kawata, H.M. Heise, Near-Infrared Spectroscopy: Principles, Instruments, Applications, John Wiley & Sons, 2001.
[18] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2004) 91–110.
[19] A.E. Abdel-Hakim, A.A. Farag, CSIFT: a SIFT descriptor with color invariant characteristics, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, pp. 1978–1983.
[20] G.J. Burghouts, J.-M. Geusebroek, Performance evaluation of local colour invariants, Comput. Vis. Image Understand. 113 (2009) 48–62.

[21] M. Brown, S. Süsstrunk, Multi-spectral SIFT for scene category recognition, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2011, pp. 177–184.
[22] B. Zhang, L. Zhang, D. Zhang, L. Shen, Directional binary code with application to polyU near-infrared face database, Pattern Recogn. Lett. 31 (2010) 2337–2344.
[23] Q. Wang, S. Li, Database of human segmented images and its application in boundary detection, IET Image Process. 6 (2012) 222–229.
[24] A. Leykin, Y. Ran, R.I. Hammoud, Thermal-visible video fusion for moving target tracking and pedestrian classification, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007.
[25] L. Zhang, L. Zhang, D. Tao, X. Huang, Tensor discriminative locality alignment for hyperspectral image spectralcspatial feature extraction, IEEE Trans. Geosci. Remote Sensing (2012) 1–15.
[26] Y. Wen, Y. Gao, S. Liu, Q. Cheng, R. Ji, Hyperspectral image classification with hypergraph modelling, in: ICIMCS, 2012, pp. 34–37.
[27] Y. Gu, C. Wang, D. You, Y. Zhang, S. Wang, Y. Zhang, Representative multiple kernel learning for classification in hyperspectral imagery, IEEE Trans. Geosci. Remote Sensing 50 (2012) 2852–2865.
[28] Multi-spectral camera, User's Manual, Ver.1.0, JAI Ltd. (October 2009).
[29] R. Achanta, F. Estrada, P. Wils, S. Süsstrunk, Salient region detection and segmentation, Comput. Vis. Syst. 5008 (2008) 66–75.
[30] Y. Zhai, M. Shah, Visual attention detection in video sequences using spatiotemporal cues, in: ACM Multimedia, 2006, pp. 815–824.
[31] R. Achanta, S. Süsstrunk, Saliency detection using maximum symmetric surround, in: International Conference on Image Processing, pp. 2653–2656.
[32] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[33] L. Zhang, M.H. Tong, T.K. Marks, H. Shan, G.W. Cottrell, SUN: a Bayesian framework for saliency using natural statistics, J. Vision 8 (2008) 1–20.
[34] Q. Wang, P. Yan, Y. Yuan, X. Li, Multi-spectral saliency detection, Pattern Recogn. Lett. 34 (2012) 34–41.
[35] D.R. Martin, C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, color, and texture cues, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2004) 530–549.
[36] S. Chatterjee, A.S. Hadi, Influential observations, high leverage points, and outliers in linear regression, Stat. Sci. 1 (1986) 379–393.
[37] G.A.F. Seber, C.J. Wild, Nonlinear Regression, Wiley-Interscience, Hoboken, NJ, 2003.