Top-Down Visual Saliency Detection in Optical Satellite Images Based on Local Adaptive Regression Kernel

Xiaoguang Cui, Yuan Tian, and Lei Ma
Institute of Automation, Chinese Academy of Sciences, Beijing, China
Email: {xiaoguang.cui, yuan.tian, lei.ma}@ia.ac.cn

***Abstract*—This paper proposes a novel top-down visual saliency detection method for optical satellite images using local adaptive regression kernels. This method provides a saliency map by measuring the likeness of image patches to a given single template image. The local adaptive regression kernel (LARK) is used as a descriptor to extract feature and compare against analogous feature from the target image. A multi-scale pyramid of the target image is constructed to cope with large-scale variations. In addition, accounting for rotation variations, the histogram of kernel orientation is employed to estimate the rotation angle of image patch, and then comparison is performed after rotating the patch by the estimated angle. Moreover, we use the bounded partial correlation (BPC) to compare features between image patches and the template so as to rapidly generate the saliency map. Experiments were performed in optical satellite images to find airplanes, and experimental results demonstrate that the proposed method is effective and robust in complex scenes.**

***Index Terms*—Visual Saliency; Top-Down; Kernel Regression; Rotation Invariance; Bounded Partial Correlation**

I. INTRODUCTION

In recent years, with the development of remote sensing technology, optical satellite images have been widely used for target detection such as harbors, cars and airplanes. High spatial resolution satellite images can provide informative details of shape, texture and context [1]. However, the data explosion of high resolution images brings more difficulties and challenges for fast image processing. Visual saliency detection aims at identifying the most significant area of interest in images rapidly so as to reduce the search space. The areas of interest can be processed with priority by the limited computing resource, thus substantially improving the efficiency of image processing [2]-[3].

Visual Saliency has been extensively studied in both the biological and computer vision literatures over the last decades. Visual Saliency is classified into bottom-up saliency and top-down saliency. Bottom-up saliency is mainly driven by low-level visual features (e.g., color, intensity and oriented filter responses) and models the data-driven visual processing in early vision. Several bottom-up models are based on the well known biological saliency model by Itti et. al. [4]. In this model, an image is decomposed into low-level feature maps across several spatial scales, and then a master saliency map is formed by linearly or non-linearly normalizing and combining these maps. Different from the biological saliency models, some bottom-up models are based on mathematical methods. For instance, Graph-based Visual Saliency (GBVS) [5] forms a bottom-up saliency map based on graph computations; Hou and Zhang [6] propose a Spectral Residual Model (SRM) by extracting the spectral residual of an image in spectral domain; Pulsed Cosine Transform (PCT)-based model [7] extends the pulsed principal component analysis to a pulsed cosine transform to generate spatial and motional saliency. Bottom-up saliency is effective for highlighting the informative regions of images. However, due to lack of top-down prior knowledge, bottom-up saliency approaches usually focus on numerous unrelated regions with low-level visual stimuli and thus miss the objects of interest. This situation is particularly common in Optical Satellite Images because backgrounds are often highly cluttered.

Compared with the bottom-up mechanism, top-down saliency models use some prior knowledge of the target to generate probability maps that are more meaningful to locate objects of interest. Recently, several top-down methods have been proposed based on learning mappings from image features to eye fixations using machine learning techniques. Zhao and Koch [8]-[9] combined saliency channels by optimal weights learned from eye-tracking dataset. Peters and Itti [10], Kienzle et al. [11] And Judd et.al. [12] learned saliency using scene gist, image patches, and a vector of features at each pixel, respectively. J. Yang et.al. [13] proposed a saliency model that jointly learns a conditional random field and a discriminative dictionary.

It is established that top-down models achieve higher accuracy than bottom-up models if enough training data is available. However, target samples are often not sufficient for training top-down models in practical applications, thus most top-down models are unreliable under such extreme condition. In this case, a top-down model which can locate saliency effectively using only few target samples is needed.

This paper uses a single template of the target to yield a scalar saliency map which indicates the statistical

© 2014 ACADEMY PUBLISHER
doi:10.4304/jmm.9.1.173-180

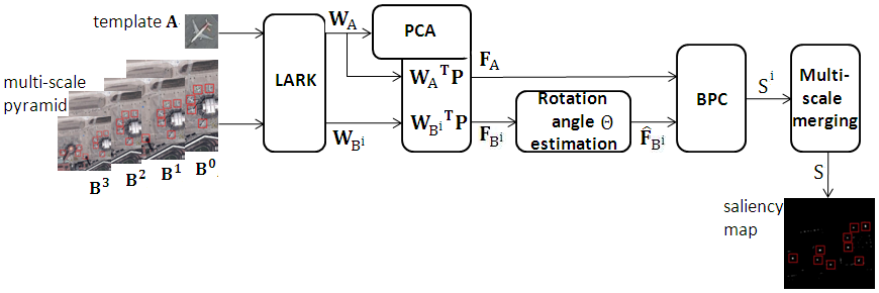


Figure 1. The framework of the proposed method

likelihood of similarity between the template and all target patches in a test image. Our method is based on the computation of local adaptive regression kernels (LARK) as descriptors from image patches, which measure the local similarity of a pixel to its neighbors both geometrically and photometrically. The LARK is an effective tool for denoising, interpolation, and deblurring in image processing [14]. Recently, the LARK has been shown to be effective in object detection [15]-[16] and bottom-up saliency detection [17] due to its invariance and robustness in various challenging conditions, such as brightness change, contrast change, low signal-to-noise ratio and even slight deformation. However, the use of LARK usually faces two thorny problems [16]: ***1)*** How to deal with the variations in scale and rotation. ***2)*** How to reduce the high computational complexity. Focusing on handling the two problems, this paper proposes the following solutions:

To solve the first issue, a multi-scale pyramid of the target image is constructed to cope with large-scale variations. On the other hand, accounting for rotation variations, histogram of kernel orientations is employed for estimating a rotation angle of image patch, and then the comparison is performed after rotating the patch by the estimated angle.

For the second issue, we use the bounded partial correlation (BPC) [19] elimination algorithm instead of the common standard full-searching (FS) algorithm to compare features between image patches and the template so as to generate the saliency map rapidly.

This paper is structured as follows. In Section II, we briefly describe the framework of the proposed approach. The details of feature extraction from the LARK descriptors are presented in Section III. In Section IV, we handle variations in scale and rotation. The BPC is introduced in Section V. Experiment and discussion is shown in Section VI, and Section VII is devoted to conclusion and future work.

II. Overview of the Proposed Approach

A graphical overview of the proposed method is shown in Fig. 1. Given a template (A) of an object of interest, we first calculate the LARK (W_A) from this template at all pixel locations, then a dimensionality reduction step using standard PCA produces a feature vector (F_A) of modest dimensions. A similar feature vector (F_{B^i}) computed from candidate patches in a multi-scale pyramid of target image (B^i) is rotated by a rotation angle Θ, which is estimated by the histogram of kernel orientations. Subsequently, the rotated feature ($\hat{F}_{B^i}$) of target image patches is compared against the template feature (F_A) using BPC to create multi-scale saliency maps (S_i). Finally, a multi-scale merging procedure is performed to generate the final saliency map (S) which shows the likelihood of similarity between the template and target image patches.

III. Feature Extraction From the LARK

The key idea of LARK is to robustly obtain local data structures by estimating distribution of gradients, and then use this structure information to determine the shape and size of a canonical kernel [14]. The LARK $K(\cdot)$ is defined as:

$$\mathrm{K}(x_i - x) = \frac{\sqrt{\det(C_i)}}{2\pi h^2} \exp\left\{-\frac{(x_i - x)^T C_i (x_i - x)}{2h^2}\right\}; \quad i = 1, \ldots p^2 \tag{1}$$

where is the spatial coordinates, p^2 is the number of pixels in a $p \times p$ local window, h is a global smoothing, C_i denotes a covariance matrix estimated from a collection of spatial gradient vectors within the local window around a position x . C_i modifies the shape of the local kernel by means of encoding the local edge structures present in the local window (See Fig. 2 for an example), and can be decomposed into three components as follows :

$$C_i = \gamma_i U_{\theta i} \Lambda_i U_{\theta i}{}^T$$
$$U_{\theta_i} = \begin{bmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{bmatrix} \tag{2}$$
$$\Lambda_i = \begin{bmatrix} \sigma_i & 0 \\ 0 & \sigma_i^{-1} \end{bmatrix}$$

where U_{θ_i} is a rotation matrix and Λ_i is a elongation matrix. γ_i and are the scaling and elongation parameters, respectively. θ_i is the kernel orientation, by which the rotation angle of image patch can be estimated in subsequent Section IV. The three parameters , and can be

© 2014 ACADEMY PUBLISHER

estimated using singular value of the spatial gradient matrix, (we refer the reader to [14] for further details).

Given a template image $A \in R^{m \times n}$ and a target image $B \in R^{M \times N}$ ($M \geq m, N \geq n$), the LARK $K^j(x_i - x)$ at a local window indexed by j is densely calculated and normalized as follows:

$$W^j{}_A(x_i - x) = \frac{K^j{}_A(x_i - x)}{\sum_{i=1}^{p^2} K^j{}_A(x_i - x)}$$
$$i = 1,...p^2; j = 1,...,mn$$
$$W^j{}_B(x_i - x) = \frac{K^j{}_B(x_i - x)}{\sum_{i=1}^{p^2} K^j{}_B(x_i - x)} \quad (3)$$
$$i = 1,...p^2; j = 1,...,MN$$

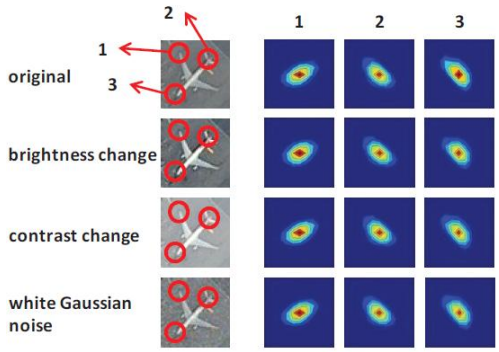


Figure 2. Examples of LARK in various challenging conditions

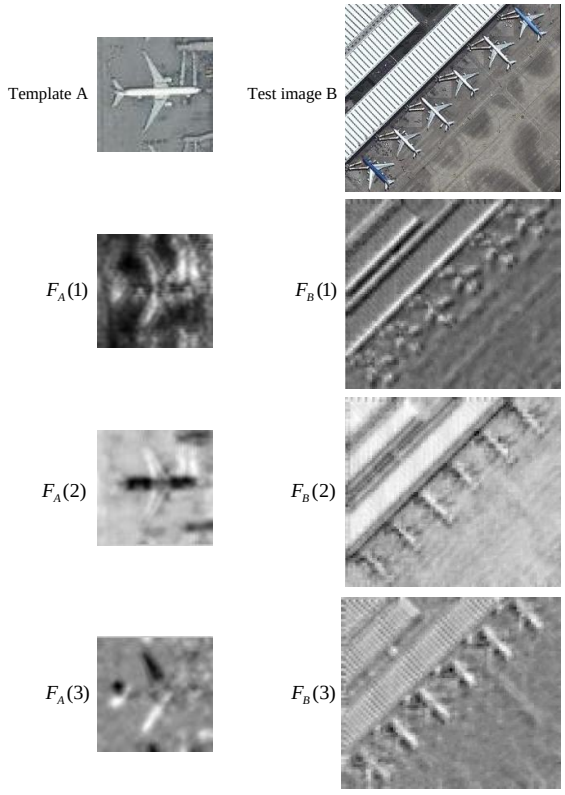


Figure 3. The illustration of feature maps F_A and F_B. The first 3 row vectors of F_A from template A and F_B from test image B are shown in the figure.

As shown in Fig. 2, the normalization of LARK brings robustness to various challenging conditions, such as brightness change, contrast change and low signal-to-noise ratio. Let W_A, W_B denote feature matrices, which are column stacked versions of $W^j{}_A, W^j{}_B$,

$$W_A = [w^1{}_A, ..., w^{mn}{}_A] \in R^{p^2 \times mn}$$
$$W_B = [w^1{}_B, ..., w^{MN}{}_B] \in R^{p^2 \times MN} \quad (4)$$

where $W^j{}_A, W^j{}_B$ are column vectors. A dimensionality reduction step using standard PCA is conducted to enhance the discriminative power and reduce computational complexity. The lower dimensional features F_A and F_B are computed by projecting W_A and W_B onto a matrix P whose columns are the first largest d principal components:

$$F_A = W_A{}^T P \in R^{mn \times d}$$
$$F_B = W_B{}^T P \in R^{MN \times d} \quad (5)$$

An example of feature maps F_A and F_B is shown in Fig. 3.

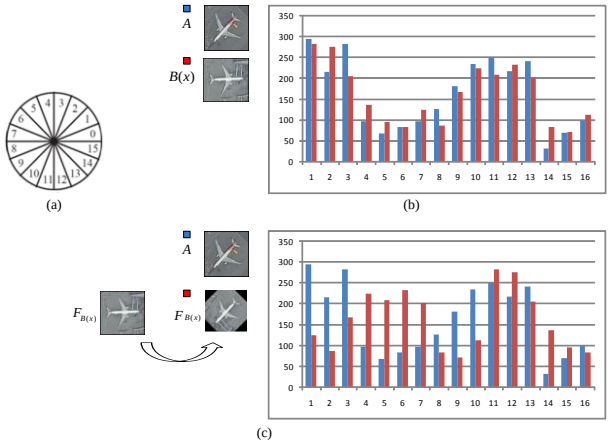


Figure 4. (a) Illustration of orientations. (b) Kernel orientation histograms of a template A and an image patch centered in the position x, B(x). (c) Histograms after rotation.

IV. Handling Variations in Scale and Rotation

A. Multi-scale Approach

We construct a multi-scale pyramid of the target image B to cope with scale variations. The first step in our method is to build the multi-scale pyramid () by a Gaussian down-sampling. Each scale of the pyramid yields a saliency map ($S^k, k = 0,...,l$) through the same processing procedures. However, the sizes of the respective saliency maps are naturally different. Therefore, we simply upscale all of the saliency maps by pixel replication so that they match the dimensions of the finest scale map B^0 (the original target image). We define $\bar{S}^k$ as the up-scaled saliency map, and then the final saliency map S is obtained by comparing the value of $\bar{S}^k$ ($k = 0,...l$) at each position x as follows:

© 2014 ACADEMY PUBLISHER

$$S_x = \max\{\bar{S}_x^{\,0}, \bar{S}_x^{\,1}, ..., \bar{S}_x^{\,k}\} \quad (6)$$

B. Rotation Invariance Approach

In order to cope with large variations in rotation, inspired by [18], we first compute the rotation angle of image patches by the histogram of the kernel orientation. Then the feature vector of the patch is rotated by the rotation angle so as to form a rotation invariance feature. In fact, this is a process of rotation alignment as shown in Fig. 4.

We use 16 orientations, which is corresponding to a sector width Δ_θ of $\pi/8$ radians, to construct the histogram. An illustration of orientations is shown in Fig. 4 (a). The r th bin of the histogram for a patch can be written as:

$$\begin{aligned} \mathrm{h}(r) &= \sum_{i,j} \delta(r - [\theta_{i,j}/\Delta_\theta]) \\ i &\in [1, p^2]; j \in [1, mn] \end{aligned} \quad (7)$$

where $\delta(\cdot)$ represents the Kronecker's delta. $[\cdot]$ denotes the rounding operator. $\theta_{i,j}$ is the kernel orientation of a pixel within a $p \times p$ local window indexed by j . The bins, corresponding to , represent the frequency of occurrence of kernel orientations. We compute the rotation angle of a patch , based on a measurement of the similarity between the histogram of the patch h_p and the template h_T . We use the city-block metric, which is equivalent to histogram intersection technique based on max-min strategy for the cases when two histograms are the same size, to check the similarity between h_p and h_T . Θ is given by

$$\begin{aligned} \Theta &= K * \Delta_\theta \\ K &= \arg\max \left\{ \sum_{r=0}^{15} \min\{h_p^k(r), h_T^k(r)\} \right\} \\ k &= 0, ..., 15 \end{aligned} \quad (8)$$

where h^k represents the histogram h shifted by bins, i.e. h^k can be calculated by

$$\mathrm{h}^k(r) = h(r + k \bmod 16) \quad (9)$$

The area below the curve, which is obtained by the intersection between h_T and the patch histogram shifted left by K bins, represents the minimum distance between and . Let $\mathrm{F}_{B(x)}$ (a chunk of F_B centered around the pixel x) be the feature vector of a patch centered in the position x, we rotate $\mathrm{F}_{B(x)}$ by Θ to form a rotation invariance feature for the purpose of eliminating the impacts of rotation variation using the following coordinate rotation transformation:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\Theta & -\sin\Theta \\ \sin\Theta & \cos\Theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (10)$$

In the following section, the likeness of an image patch to the template is measured between the computed features, $\hat{F}_{B(x)}$ and F_A .

V. Feature Comparison Using the Bounded Partial Correlation

The LARK is designed with detection accuracy as a high priority, therefore an improvement of the computational complexity is required for practical applications. Toward this end, we could benefit from an efficient searching method [19], referred to as bounded partial correlation (BPC), based on the normalized correlation (NC). The method consists in checking at each position a suitable elimination condition relying on the evaluation of an upper-bound for the NC. The positions that dissimilate the template are rapidly skipped by the check, hence the computational complexity of feature comparison can be reduced.

Let (x, y) denote the current patch position in the target image. The NC at denoted by $\eta(x, y)$, is defined as:

$$\eta(x, y) = \frac{\Psi(x, y)}{\left\|\hat{F}_{B(x,y)}\right\| \cdot \|F_A\|} \quad (11)$$

where represents the correlation between the template and the current patch:

$$\psi(x, y) = \hat{F}_{B(x,y)}^T F_A \quad (12)$$

and $\|\cdot\|$ is the l^2 -norm operator. $\left\|\hat{F}_{B(x,y)}\right\|$ can be obtained efficiently by using box-filtering techniques [18], and $\|F_A\|$ can be computed at the beginning. The remaining challenge is to compute the numerator. The BPC derives upper bounds for the correlation, based on the Jensen and Cauchy-Schwarz inequalities, to skip search points without calculating .

Let be an upper bound of , i.e. . We can obtain an upper bound of the NC by normalizing :

$$\frac{\beta(x, y)}{\left\|\hat{F}_{B(x,y)}\right\| \cdot \|F_A\|} \geq \frac{\psi(x, y)}{\left\|\hat{F}_{B(x,y)}\right\| \cdot \|F_A\|} = \eta(x, y) \quad (13)$$

We use a threshold τ to decide whether the current patch is sufficiently similar to the template, if the below inequality holds, the comparing process at position (x, y) should be skipped.

$$\frac{\beta(x, y)}{\left\|\hat{F}_{B(x,y)}\right\| \cdot \|F_A\|} < \tau \quad (14)$$

$\beta(x, y)$ is initially obtained based on the Jensen inequality, and then a tighter bound is derived by deploying the Cauchy-Schwarz inequality. The upper bound adopting the partial correlation associated with row τ can be expressed as:

© 2014 ACADEMY PUBLISHER

$$\beta(x,y)=\sum_{j=0}^{r}\sum_{i=0}^{m}\hat{F}^{T}_{B(x,y)}(x+i,y+j)\cdot F_A(i,j)+\sqrt{\sum_{j=r+1}^{n}\sum_{i=0}^{m}\left\|\hat{F}^{T}_{B(x,y)}(x+i,y+j)\right\|^2}\sqrt{\sum_{j=r+1}^{n}\sum_{i=0}^{m}\left\|F_A(x+i,y+j)\right\|^2} \quad (15)$$

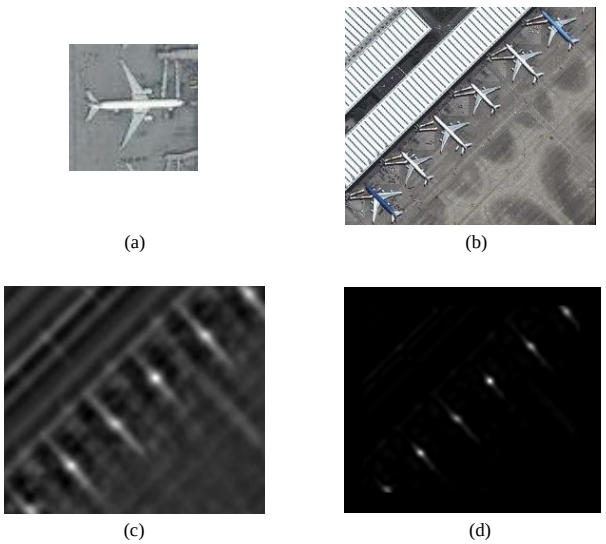
where the first term computed only a particular portion of the actual correlation function, referred to as a partial correlation, and the second term derived from the application of the Cauchy-Schwarz inequality bounds the residual portion of the correlation function.

It is clear that, with the increasing of r, the bounding function gets closer to the actual correlation and the elimination condition becomes more effective. However, the computational savings will reduce with r because of the increasingly computation of the first term in Equation. 15. Hence, r turns out to be the fundamental parameter of the BPC, its choice implying a trade-off between the elimination efficiency and the computational complexity. As shown in Section VI, choosing proper r can yield a considerable speedup with respect to the standard exhaustive search algorithm.

After the NC is calculated, it can be directly used as a measure of similarity between the template and target image patches. However, in order to make the saliency map provides better contrast and dynamic range, we take the canonical correlation value as the final test statistic comprising the values in a saliency map:

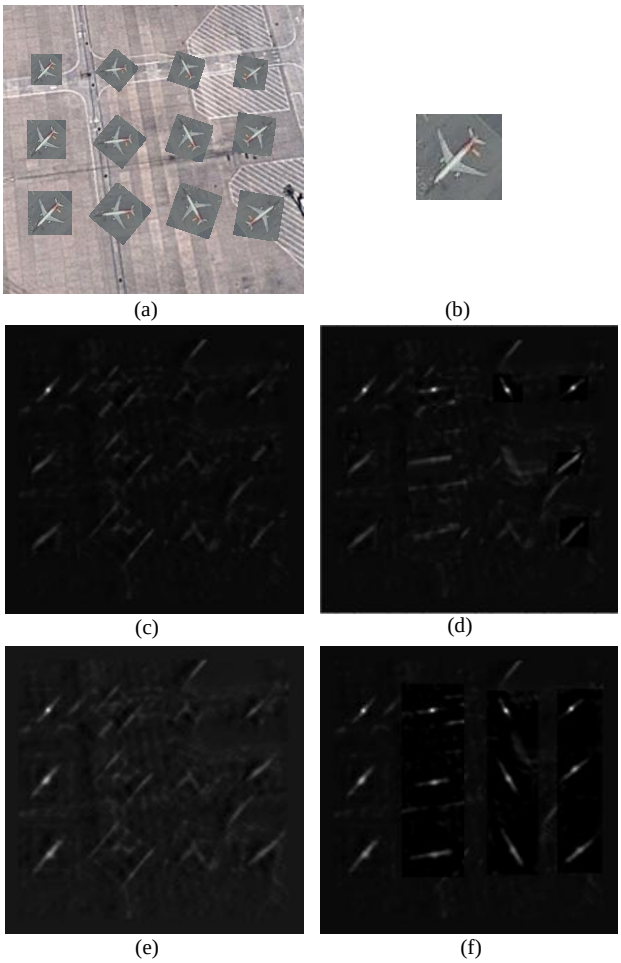
$$S(x,y)=\begin{cases}0 & \text{the sikpped postitions}\\ \dfrac{\eta^2(x,y)}{1-\eta^2(x,y)} & \text{other wises}\end{cases} \quad (16)$$
$$i=1,...,l$$

Fig. 5 shows an example of saliency map based on the NC and the canonical correlation. Obviously, the saliency regions are more prominent in the map generated by the canonical correlation.



(a) (b) (c) (d)

Figure 5. An example of saliency map based on the NC and the canonical correlation.(a) template (b) test image (c) saliency map with NC (d) saliency map with canonical correlation.



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

Figure 6. Detection results of the original LARK (c), LARK with rotation invariance approach(d), LARK with multi-scales approach(e) and the proposed model(f), performed on a simulation image(a) which contains targets with various scales and rotation. (b) is the target template.

VI. Experiment and Discussion

To verify the effectiveness of our model, the experiment for detecting planes was performed on the real optical satellite images. The test dataset includes 20 gray-scale images with the size of 750×750, all from Google Earth. A total of 367 targets are involved in the experiment.

A. The Performance of the LARK

We computed LARK of size 9×9 as described, then the 81-dimensional local feature was reduced to 4-d (contains about 90% largest principal components) by PCA. A multi-scale pyramid with five scale factors 1.0, 0.8, 0.6, 0.4 and 0.2 was constructed for target images to deal with scale variations in the beginning.

We first demonstrate the performance of the proposed model on handling variations in scale and rotation. The original LARK without processing for variations was implemented for bench-marking against the proposed model. Firstly, the efficiency of the original LARK and the proposed model were examined. Then the LARK with multi-scales and with rotation invariance approach were examined, respectively. The results are shown in Fig. 6 ,

© 2014 ACADEMY PUBLISHER

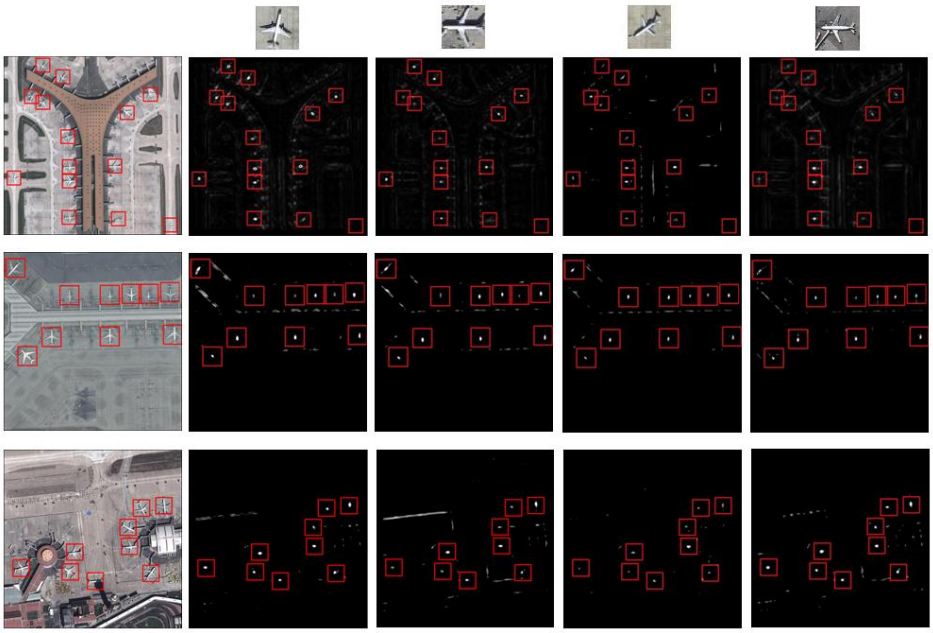


Figure 7. Detection results using different templates. The airplane locations are marked with a red rectangle.

and the Receiver Operation Characteristic (ROC) curves in Fig. 8 represent an accurate comparison. It is obvious that both the multi-scales and rotation invariance approach results in a higher accuracy, and the proposed model is much more robust to variations in scale and rotation than the original LARK under complex scenes.

Next we illustrate the impact of target template on our model. Fig. 7 shows some results on the test dataset using different target templates (51×51). It can be seen that the performance of our model is not seriously affected by a choice of target template, even though the templates are contaminated by black shadow or airport building, the results are still acceptable. For the sake of completeness, we show the ROC curves using four different target templates in Fig. 7. The ROC curves illustrate that about 80\% detection rate is achieved with a handful of false positives.

B. The Efficiency Evaluation of the BPC

In the BPC process, the threshold τ is set to 0.5, thus positions with an upper bounded NC lower than 0.5 are skipped in feature comparison. To illustrate the impact of the parameter r on execution time, we define a skip ratio R_s as:

$$R_s = \frac{S}{W \times H} \tag{17}$$

where S is the number of skipped pixels, W, H denote the width and height of the image, respectively. The relationship between R_s and r is shown in Fig. 10. It can be seen that R_s increases nearly linearly with r. Obviously, the elimination condition becomes more effective with the increasing of r, leading more pixels to be skipped. However, the execution time is not always reduced with R_s. As shown in Fig. 11, it is reduced with R_s at first and then turns back to a high level. The reason for this phenomenon is that, with the increasing of R_s, the computation of the elimination condition equation becomes so large that the computational savings by BPC can be neglected.

To further examine the performance of BPC, we compare the feature comparison execution time of BPC with the standard full-searching (FS) algorithm. It can be seen in TABLE I that, as discussed above, the computatio n would be heavy if the parameter r is too large or too small. Hence, selecting a suitable r is necessary to yield a considerable speedup. Our experiments indicated that a rather conservative choice, satisfies $r/n = 0.3$, would be favorable. The computing speed of BPC with $r = 15$ is over three times faster than FS.

TABLE I. THE AVERAGE FEATURE COMPARISON EXECUTION TIME OF ALL TEST IMAGES USING THE BPC AND FS

	average execution time(s)
FS	32.67
BPC with r = 5	29.41
BPC with r = 10	17.85
BPC with r = 15	9.71
BPC with r = 25	15.36
BPC with r = 30	19.67
BPC with r = 40	27.31
BPC with r = 45	30.25

VII. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel top-down saliency detection method by comparing the similarity between images patches and a single given target template based on local adaptive regression kernels. Our main contribution is that an effective approach is provided to deal with variations in scale and rotation. In addition, the computation complexity of the feature comparison process is significantly reduced by introducing the BPC technology. Experiment results performed in optical

© 2014 ACADEMY PUBLISHER

satellite images indicate that our method is robust and efficient for saliency detection in complex scenes.

It should be noted that although the computing speed of the feature comparison is significantly improved by the benefit of BPC, the computation complexity in feature extraction is still heavy. Therefore, improvement of the computational complexity of feature extraction from LARK is a direction of future research worth exploring.

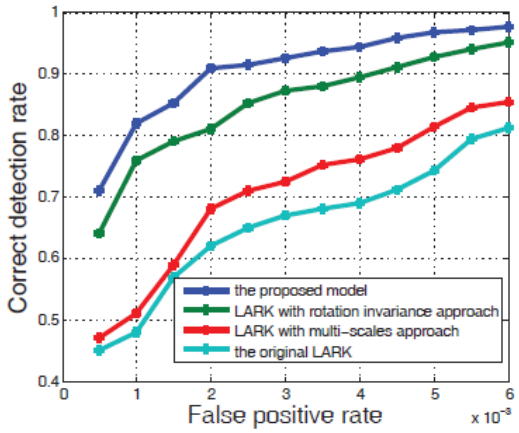


Figure 8. ROC curves of the proposed model, LARK with multi-scales approach, LARK with rotation invariance approach and the original LARK, performed on the test dataset (contains 20 real optical satellite images).

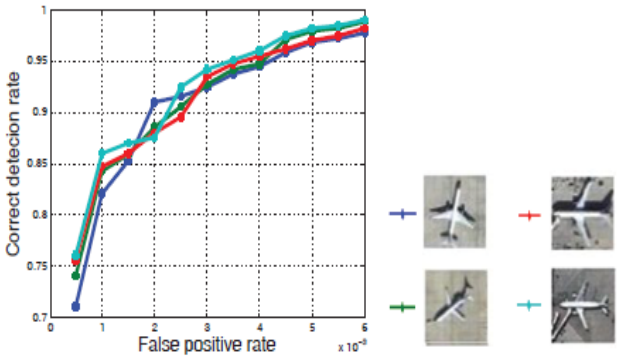


Figure 9. ROC curves using different templates performed on the test dataset.

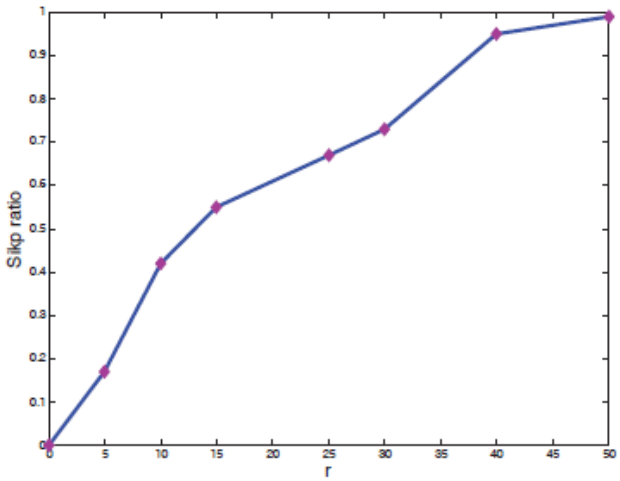


Figure 10. Skip ratio with various r. The skip ratio is average value calculate using all test images.

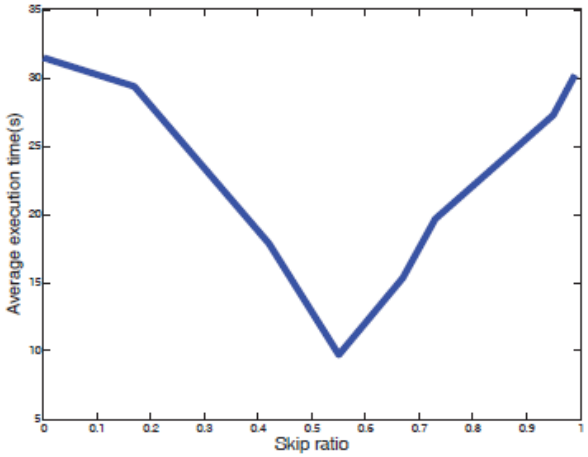


Figure 11. The relationship between the average execution time and the skip ratio (performed on a 2.80GHz Intel Core2 PC using the program Visual C++, the size of target images and template are 750* 750 and 51*51, respectively)

ACKNOWLEDGMENT

This work was supported by the National Science Foundation of China No. 61203239, No. 61005067 and No. 61101222.

REFERENCES

[1] M. Li, L. Xu, and M. Tang, "An extraction method for water body of remote sensing image based on oscillatory network," *Journal of multimedia*, vol. 6, no. 3, pp. 252–260, 2011.

[2] Q. Zhang, G. Gu, and H. Xiao, "Image segmentation based on visual attention mechanism," *Journal of multimedia*, vol. 4, no. 6, pp. 363–369, 2009.

[3] B. Yang, Z. Zhang, and X. Wang, "Visual important-driven interactive rendering of 3d geometry model over lossy wlan," *Journal of networks*, vol. 6, no. 11, pp. 1594–1601, 2011.

[4] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 20, pp. 1254–1259, 1998.

[5] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *in Advances in Neural Information Processing Systems*, 2007, pp. 542–552.

[6] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," *in IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[7] Y. Yu, B.Wang, and L.Zhang, "Bottom-up attention: Pulsed pca transform and pulsed cosine transform," *Cognitive eurodynamics*, vol. 5, no. 4, pp. 321–332, 2011.

[8] Q. Zhao and C. Koch, "Learning a saliency map using fixated locations in natural scenes," *Journal of Vision*, vol. 11, no. 3, pp. 1–15, 2011.

[9] --, "Learning visual saliency," in Information Sciences and Systems Conference, 2011, pp. 1–6.

[10] R. Peters and L. Itti, "Beyond bottom-up: Incorporating task dependent influences into a computational model of spatial attention," *in IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, pp.1–8.

© 2014 ACADEMY PUBLISHER

[11] B. Scholkopf, J. Platt, and T. Hofmann, "A nonparametric approach to bottom-up visual saliency," *in Advances in Neural Information Processing Systems*, 2007, pp. 689–696.
[12] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," *in International Conference on Computer Vision*, 2009, pp. 2106–2113.
[13] J. Yang and M.-H. Yang, "Top-down visual saliency via joint crf and dictionary learning," *in IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1–8.
[14] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 349–366, 2007.
[15] H. J. Seo and P. Milanfar, "Using local regression kernels for statistical object detection," *in IEEE International Conference on Image Processing*, 2007, pp. 2380–2383.
[16] --, "Training-free, generic object detection using locally adaptive regression kernels," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1688–1704, 2010.
[17] --, "Static and space-time visual saliency detection by self-resemblance," *Journal of Vision*, vol. 9, no. 12, pp. 1–27, 2009.
[18] S. Mattoccia, F. Tombari, and L. D. Stefano, "Fast fullsearch equivalent template matching by enhanced bounded correlation," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 528–538, 2008.
[19] F. Ullah and S. Kaneko, "Using orientation codes for rotation-invariant template matching," *Pattern Recognition*, vol. 37, pp. 201–209, 2004.

XiaoGuang Cui Ph. D. candidate at the Institute of Automation, Chinese Academy of Sciences. He received his master degree from Southwest University of Science and Technology, China, in 2010. His research interest covers remote sensing image processing, anomaly detection and visual saliency detection. Corresponding author of this paper.

Yuan Tian Professor at the Institute of Automation, Chinese Academy of Sciences. He received his Ph. D. degree from Xi′an Jiaotong University in 1998. From 1999 to 2001, he worked as a postdoctoral research fellow at the Institute of Automation, Chinese Academy of Sciences. His research interest covers image processing, pattern recognition, computer vision, and intelligent visual surveillance.

Lei Ma Research assistant at the Institute of Automation, Chinese Academy of Sciences. He received his Ph. D. degree from Institute of Automation, Chinese Academy of Sciences, Beijing, in 2012. His research interest covers remote sensing image processing and target recognition.

© 2014 ACADEMY PUBLISHER