

Research Article

A Method of Waypoint Selection in Aerial Images for Vision Navigation

Lin Song, Yong-mei Cheng, Lu Yu, and Liang Yu

School of Automation, Northwestern Polytechnical University, Xi'an 710072, China

Correspondence should be addressed to Lin Song; linsong0818@gmail.com

Received 1 June 2014; Revised 26 August 2014; Accepted 14 September 2014; Published 8 October 2014

Academic Editor: Nicusor Iftimia

Copyright © 2014 Lin Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a novel method to select waypoints from aerial images of candidate flying regions via matching suitability analysis, which is based on visual attention mechanism and feature attribute classification. At first, visual attention mechanism is used to get the saliency map of the initial image by low-rank recovery and sparse coding. The salient regions are selected to be as preparatory results with threshold constraint and nonmaxima suppression. Then we use support vector machine (SVM) to divide the preparatory results into two classes for suitable or unsuitable waypoints based on their feature attributes, which can be represented by two edge-based descriptors and two correlation-based descriptors. The experimental results show that the proposed method is valid and effective.

1. Introduction

In recent years, the research on vision navigation based on scene matching technique has attracted more and more attentions for its accuracy and independence [1]. In order to assure the matching precision, it is a primary step to select waypoints from aerial images of candidate flying regions for scene matching, which can be implemented via suitability analysis [2–4].

Some researchers have made efforts to solve the problem of suitability analysis. Yang et al. [2] used gray-based descriptors and edge-based descriptors as the input of SVM to classify the matching suitability of the image. Jiang and Chen [5] provided a hierarchical way of selecting optimal scene matching areas. Liu et al. [6] presented the method of selecting matching areas using independent pixel numbers and variance. In [7] it was suggested that information entropy and summation of image gradient can be used for evaluating image suitability. However, features used in the above methods are inadequate to describe the suitability and there is redundancy information among the feature descriptors. Research has shown that the suitability is considered to consist of information, obviousness, stability, and uniqueness [1]. From the view of visual analysis, the information and obviousness can be analyzed via visual saliency. Visual

saliency is a cognitive procedure which can rapidly select a small set of highly informative or visually salient objects from a scene for further processing [8]. The stability and uniqueness are the feature attributes of an image. So we can transform the suitability analysis to the combination of visual saliency analysis and feature attributes classification.

For visual saliency analysis, in order to avoid the uncertainty influence of statistic features, Li et al. [9] proposed the visual saliency analysis method based on length of incremental coding, which is based on sparse coding [10, 11] and center-surround model (C-S); Han et al. [12] proposed saliency analysis method based on weighted length of incremental coding. They are all efficient approaches to select regions with local saliency, but for matching suitability the saliency should be with unique structure information, which means it is sparse in the global image. So low-rank recovery is introduced to analyze the global and local saliency with sparse coding for preparatory selection. For feature attributes classification, SVM is used to analyze stability and uniqueness for optimizing selection. So this paper presents a practical framework for waypoint selection, as illustrated in Figure 1.

The proposed framework consists of two major components: a preparatory selection model based on visual saliency analysis and an optimal selection model based on SVM. In the formal model, the initial image is decomposed as the sum

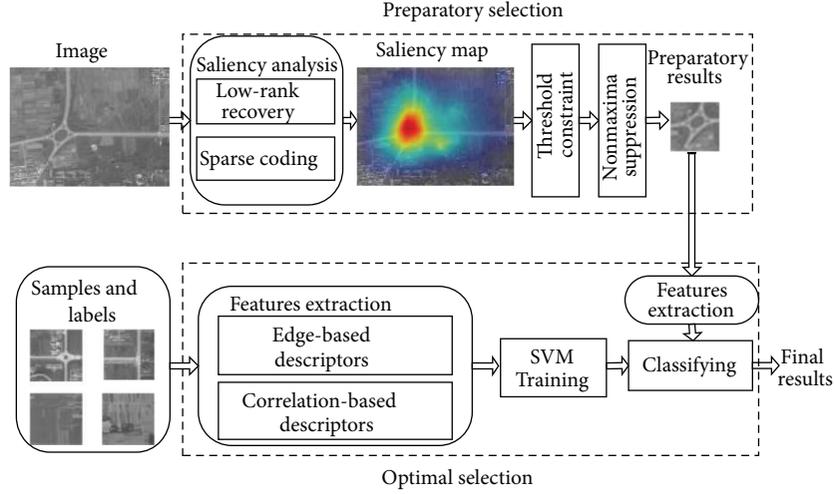


FIGURE 1: Framework of our proposed method.

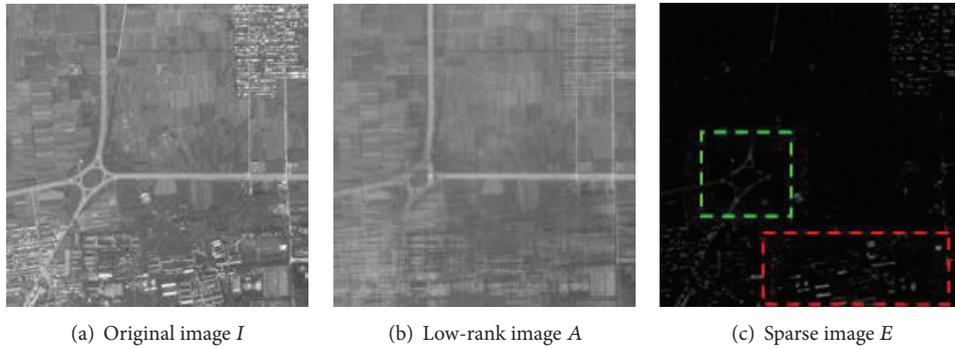


FIGURE 2: Low-rank matrix recovery.

of sparse matrix and low-rank matrix, and then saliency of sparse matrix and low-rank matrix is analyzed, respectively, to construct a new saliency map. The preparatory selection results are got with threshold constraint and nonmaxima suppression. In the second component, SVM is used for optimizing selection, and the input vector is composed of four measure parameters based on the edge and cross correlation surface.

The rest of the paper is organized as follows. Section 2 describes the preparatory selection model based on visual saliency analysis. Section 3 presents optimal selection of waypoints model based on feature attribute classification. Section 4 reports and analyzes evaluation results. Finally, conclusions are drawn in Section 5.

2. The Preparatory Selection Based on Visual Saliency Analysis

Salient objects can be viewed as a small number of foregrounds, which are different from the surrounding backgrounds. So low-rank matrix recovery is introduced to separate foreground from background [13]. The low-rank

matrix A and the sparse matrix E are recovered by optimizing the constraint,

$$\min_{A,E} (\text{rank}(A) + \lambda \|E\|_0), \quad \text{s.t. } I = A + E, \quad (1)$$

where $\text{rank}(\cdot)$ represents matrix rank, $\|\cdot\|_0$ represents l_0 -norm, and λ is the weight parameter which is used to weight the sparse relationship between the rank of A and the sparsity of E . Given a suitable λ , we want to be able to get a pair of (A, E) . However, it is a nonconvex problem for the existence of $\text{rank}(\cdot)$ and $\|\cdot\|_0$. Usually, (1) is relaxed by solving the convex optimization problem

$$\min_{A,E} (\|A\|_* + \lambda \|E\|_1), \quad \text{s.t. } I = A + E, \quad (2)$$

where $\|A\|_*$ is the sum of singular values of matrix A and $\|\cdot\|_1$ represents l_1 -norm. The optimization problem (2) is solved by utilizing augmented Lagrange multiplier (ALM) to get A and E , respectively, as shown in Figures 2(b) and 2(c).

There are two kinds of saliency in aerial images. One is with its unique information of scene structure (shown as green dotted line in Figure 2(c)) and the other is some small man-made areas with high brightness where there is no structure information (shown as red dotted line in

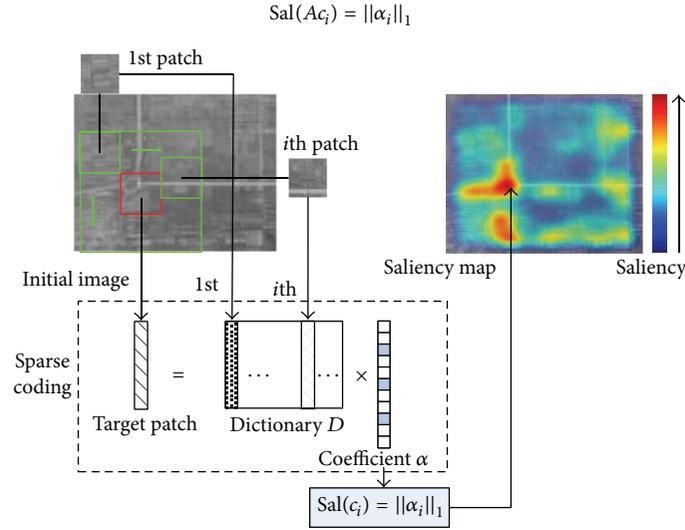


FIGURE 3: Local saliency based on sparse coding.

Figure 2(c)). We can see that the second saliency is the same as its surrounding objects in the image A whereas the first saliency in both A and E is different from its surrounding objects. So we can separate the two kinds of saliency by local saliency analysis of A and E , respectively, based on center-surround model (C-S).

Here sparse coding is introduced to describe the local saliency with C-S model. Sparse coding codes the center patch over a dictionary constructed by the surrounding patches. If the center patch is similar to its surroundings, it has sparse coefficients. Denote by A_{c_i} the i th patch of A , and $D(A_{c_i})$ is dictionary consisting of surrounding patches, which is represented by a set of vectors $D(c_i) = \{d_i^1, d_i^2, \dots, d_i^m\}$ as a dictionary. Consider $A_{c_i} \cap d_i^j = \emptyset$. So the problem of saliency based on sparse coding is shown as follows:

$$\alpha_i = \arg \min_{\alpha_i} \{ \|A_{c_i} - D(A_{c_i}) \cdot \alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \}, \quad (3)$$

where λ is the balance factor between sparsity and data integrity. The l_1 -norm optimization problem can be solved efficiently by Lasso method [11]. The local saliency of image patch A_{c_i} is obtained as (4). The process is shown as in Figure 3. Consider

$$\text{Sal}(A_{c_i}) = \|\alpha_i\|_1. \quad (4)$$

All patches of the image can be calculated, so we can get the saliency map $\text{Sal}(I)$.

We calculated the local saliency in both A and E based on sparse coding. And then the new saliency could be represented by the following function:

$$\text{NSal}(I) = \text{Sal}(A) \times \text{Sal}(E). \quad (5)$$

A certain threshold T is used to judge possible waypoints, and the rule is

$$P_i(I) = \begin{cases} 1 & \text{NSal}(I_{c_i}) > T \\ 0 & \text{others.} \end{cases} \quad (6)$$

For the region $P_i(I) = 1$, the nonmaxima suppression is used to get M peaks in $\text{NSal}(I)$, which are the centers of possible waypoints.

3. Optimal Selection Based on Feature Attribute Classification

It is a problem to judge whether there are stability and uniqueness of preparatory results. From the viewpoint of pattern recognition, it can be solved by two-class classification. So, SVM is introduced.

3.1. Introduction of SVM. SVM lies in strong connection to the underlying statistical learning theory, where it implements the structural risk minimization for solving the problem of two-class classification [2]. SVM has advantages in solving the problems like small samples, high dimensions, and large scale.

Given a training sample set $D = \{(x_i, y_i)\}_{i=1}^N$ of labeled examples, with each input $x_i \in R^n$ and the output label $y_i \in \{-1, 1\}$, N is the number of training samples. The best hyperplane $w^T x + b = 0$ (b is a constant) to separate two classes is achieved by satisfying the constraint:

$$\max_{w, b} \frac{1}{\|w\|}, \quad \text{s.t. } y_i (w^T x_i + b) \geq 1, \quad i = 1, \dots, N, \quad (7)$$

where $\|\cdot\|$ is l_2 -norm of a vector, so (7) is equivalent to minimize $(1/2)\|w\|^2$ with the same constraint. The decision function is

$$y(x) = \text{sign} [w^T x_i + b]. \quad (8)$$

When the original set will not be linearly separable, it is common to define a soft margin by including variables ξ_i and parameter $c > 0$

$$\min_{w,b,\xi} \left(\frac{1}{2} w^T w + c \sum_{i=1}^N \xi_i \right), \quad \text{s.t. } y_i [w^T x_i + b] \geq 1 - \xi_i. \quad (9)$$

The original sample can be mapped into a high-dimensional space F (named feature space) by nonlinear transform $\Phi : R^n \rightarrow F$ and w is expressed as the dual form $w = \sum_{i=1}^N \alpha_i y_i x_i$. So classification output can be predicted using the decision function, as

$$\begin{aligned} y(x) &= \text{sign} \left[\sum_{i=1}^N \alpha_i y_i \Phi(x) \Phi(x_i) + b \right] \\ &= \text{sign} \left[\sum_{i=1}^N \alpha_i y_i K(x, x_i) + b \right], \end{aligned} \quad (10)$$

where $K(x_i, y_j)$ is the kernel function. There are three forms of kernel function: radial basis function and the linear and polynomial kernel. Preliminary researches suggest that the radial basis function outperforms the others. So the radial basis function kernel in the following equation will be used in the classification:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0. \quad (11)$$

3.2. Feature Selection. Suitable descriptors as the input vector of SVM can optimize computational efficiency and gain the better classification results. Here measure parameters based on the edge and cross correlation surface are considered for stability and uniqueness analysis.

3.2.1. Stability. The stability is an important feature attribute of an aerial image, which is suitable for scene matching. So we need to select waypoints with stable features. Edge complexity and edge density are selected to be measure parameters of stability.

(1) Edge Complexity. Edge complexity is a homogeneity parameter of edge texture distribution. When it is smaller, the image is more smoother, which will lead to mismatching more easily. Edge complexity is calculated by

$$EC(x, y) = \frac{\sum_{(m,n) \in \Gamma(x,y)} I_x(m, n)}{\sum_{(m,n) \in \Gamma(x,y)} I_{xy}(m, n)}, \quad (12)$$

where $\Gamma(x, y)$ is a local neighborhood with the center (x, y) and $I_x(m, n)$ and $I_{xy}(m, n)$ are 1-dimensional derivative and 2-dimensional derivative, respectively.

(2) Edge Density. Edge density can show the concentration of features in the original image. It is computed by

$$ED(x, y) = \frac{N_{\text{Edge Pixel}}(x, y)}{N(x, y)}, \quad (13)$$

where $N_{\text{Edge Pixel}}(x, y)$ is the number of points in the neighborhood with the center (x, y) . $N(x, y)$ is the number of pixels in the neighborhood.

3.2.2. Uniqueness. The global uniqueness of waypoints is analyzed to avoid the selection of repeated scenery areas. The uniqueness is determined by cross correlation plane statistic feature. Cross correlation plane R is computed pixel by pixel in the whole image C via matching waypoint image W . We use two features of cross correlation, Submaxratio and Ngb8maxratio:

$$\begin{aligned} R(x, y) &= \left(\sum_{u=1}^n \sum_{v=1}^m (W(u, v) - \bar{W})(C(x+u, y+v) - \overline{C_{xy}}) \right) \\ &\quad \times \left(\sqrt{\sum_{u=1}^n \sum_{v=1}^m (W(u, v) - \bar{W})^2} \right. \\ &\quad \left. \times \sqrt{\sum_{u=1}^n \sum_{v=1}^m (C(x+u, y+v) - \overline{C_{xy}})^2} \right)^{-1}, \end{aligned} \quad (14)$$

where \bar{W} is the mean of W and $\overline{C_{xy}}$ is the mean of an area with the same size $m \times n$ as W in C .

(1) Submaxratio. Submaxratio denotes ratio of secondary maximum peak to maximum peak, which is computed using

$$\text{SMR} = \frac{V_{\text{sub}}}{V_{\text{max}}}, \quad (15)$$

where V_{sub} is secondary high correlation peak and V_{max} is maximal correlation peak. It means the waypoint has better uniqueness when the value of SMR is closer to zero.

(2) Ngb8maxratio. Ngb8maxratio represents ratio of maximum of eight neighbor peaks to maximum peak, which is computed using

$$\text{NMR} = \frac{V_{\text{ngb}}}{V_{\text{max}}}, \quad (16)$$

where V_{ngb} is maximum of eight neighbor peaks and V_{max} is maximal correlation peak.

3.3. Training and Classifying. Select 100 sample images as a training set and label each image manually. 50 images are waypoints, and the other 50 images are nonwaypoints, which are shown as in Figure 4.

The image is decrypted by the measure parameter vector based on the edge and cross correlation surface. So the vector $\text{In} = [\text{EC}, \text{ED}, \text{SMR}, \text{NMR}]$ is used as the input for SVM. The feature vectors should be normalized before training. The best parameters $c = 6.8$ and $\gamma = 0.0769$ for the SVM classifying using (11) are obtained via training. In the testing, each of preparatory results is decrypted as In and normalized, which is put into SVM for classifying suitable or unsuitable results.

4. Experiments

As known at present, automatic selection method of waypoints has not been reported up to now, so there is no public

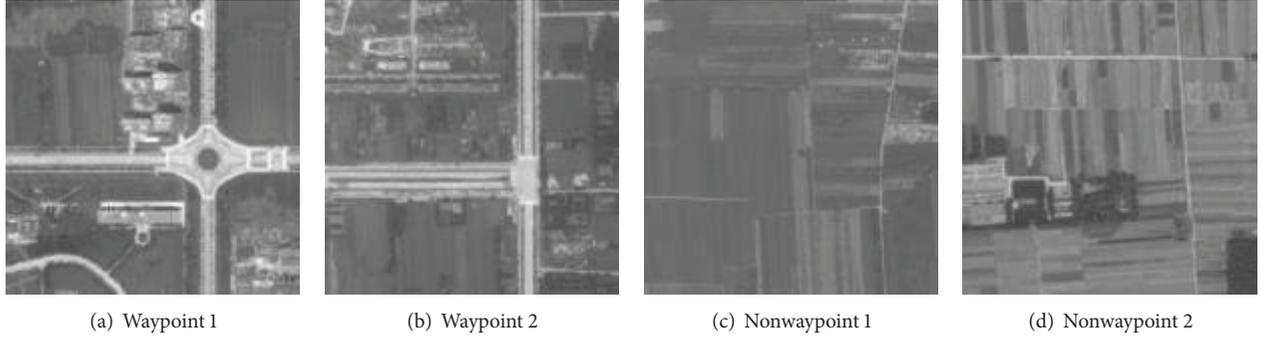


FIGURE 4: Training samples.

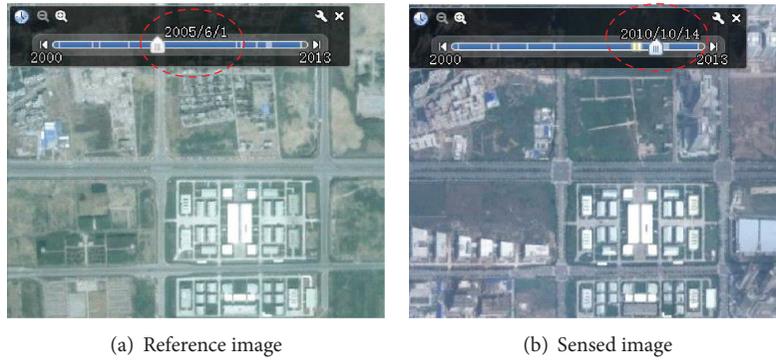


FIGURE 5: Data source from Google Earth.

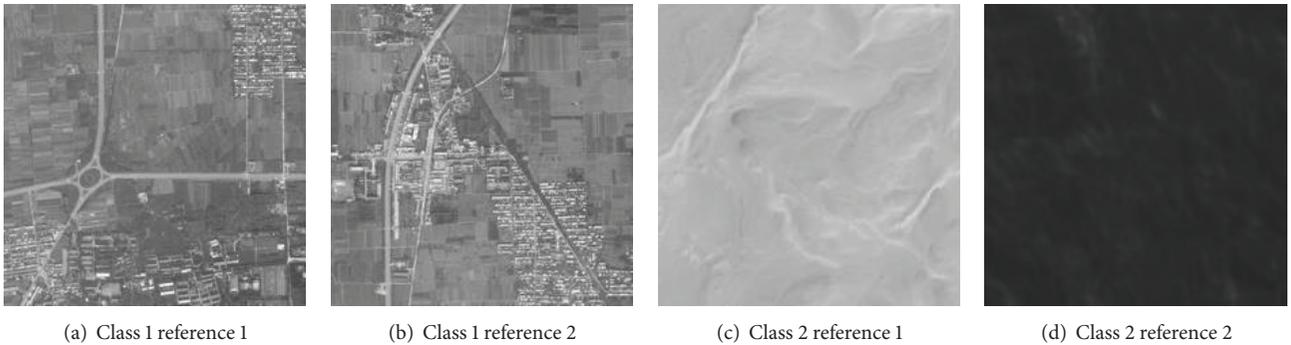


FIGURE 6: Reference images.

dataset for method validation. To evaluate the performance of our method, experiments on aerial images from Google Earth are conducted. The image pair of the same scene is taken at different time, as shown in Figure 5. One is used as reference image for waypoint selection, and the other is as sensed image used to verify the suitability of selected waypoints. For simplicity, the size of reference image and sensed image is set to be 223×223 and the size of waypoint image is set to be 51×51 in all experiments.

There are two kinds of reference images. One is called Class 1 reference, part of which is with its unique information of scene structure and can be selected as waypoints. The other is called Class 2 reference which is without any unique information of scene structure at all, as shown in Figure 6. The quantities of the two kinds are 150, respectively.

TABLE 1: Result of preparatory selection.

Reference name	Image quantity	Waypoints selected
Class 1	150	374
Class 2	150	210

4.1. Preparatory Selection. The method based on sparse coding and low-rank matrix recovery is used to implement preparatory selection of waypoints. The result is shown as in Table 1.

A few waypoints can be selected from one reference image, so the quantity of waypoints is larger than the quantity of references. Part results of Class 1 are shown as in Figure 7. In order to reduce the complexity of analysis,

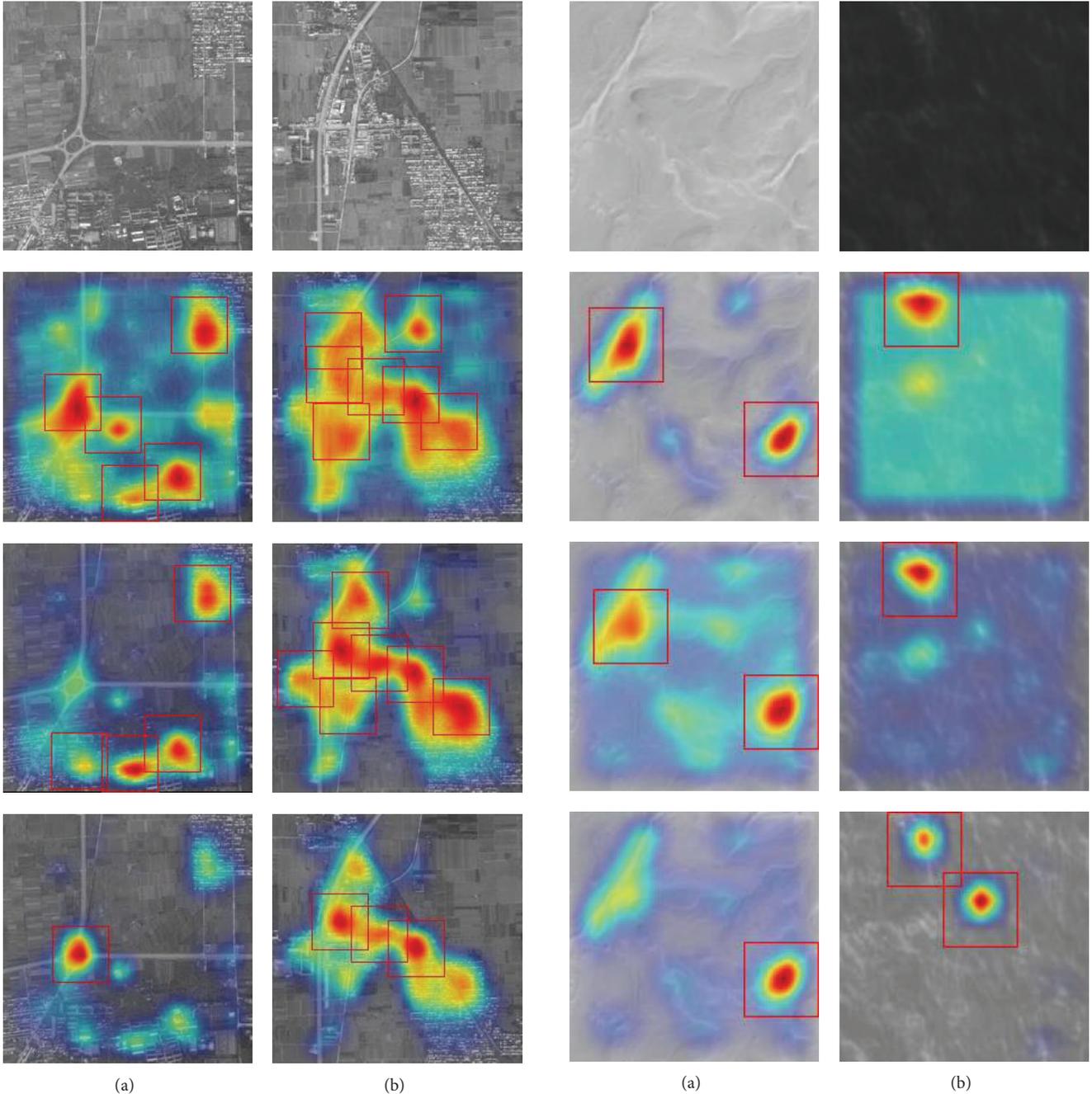


FIGURE 7: Results of waypoints in Class 1 (the first line is two reference images; the second line is the results of Li et al. [9]; the third line is Han et al.'s [12]; the fourth line is ours).

saliency coefficient is normalized between 0 and 1. Set $\text{Threshold}_{\text{saliency}} = 0.75$, and the regions are selected as waypoints, which of local saliency maximum in the centre are higher than $\text{Threshold}_{\text{saliency}}$. For comparison, two state-of-the-art saliency analysis methods are introduced: Li et al.'s [9] and Han et al.'s [12]. The results of waypoints selection are shown in Figure 7.

From Figure 7 we can see that Li et al.'s [9] and Han et al.'s [12] saliency emphasized the areas with more brightness information, most of which are without unique information

FIGURE 8: Results of waypoints in Class 2 (the first line is two reference images; the second line is the results of Li et al. [9]; the third line is Han et al.'s [12]; the fourth line is ours).

of scene structure. Our method can extract the areas with salient structure information and effectively inhibit the disturbance of brightness; for example, in the fourth line results, just the traffic intersections are extracted because of their structure information. Therefore, the number of the waypoint candidates is less than the results of the former methods.

When we analyzed the reference images from Class 2 in the preparatory selection, there were still saliency regions in the images of Class 2, as shown in Figure 8.

TABLE 2: Results of optimal selection by SVM.

Reference	Optimal selection	Matching results	
		Correct	False
Class 1	Waypoint	281	4
	Nonwaypoint	59	30
Class 2	Waypoint	0	2
	Nonwaypoint	0	208

We note that the methods are ineffective in analyzing the saliency in the references from Class 2. It is because of the normalization of the saliency coefficients. The results in Class 2 are not suitable to be waypoints, so we need to classify the results in Class 1 and the results in Class 2 with SVM.

4.2. Optimal Selection. SVM is used to optimally select waypoints by classifying the feature attributes. To evaluate the results of classification, cross correlation matching method is used for verification (it is thought to be a correct matching when the matching error is smaller than 5 pixels). The result is shown as in Table 2.

There are two kinds of mistakes in the process of classification. One is called “undetected,” which means a waypoint is classified as nonwaypoint. The other is called “false detected,” which means a nonwaypoint is classified as waypoint. The former error is tolerable, but the latter is fatal for vision navigation. So the latter should be avoided or be reduced as much as possible. The classification rate is shown as in Table 3.

From Table 3, we know that though there are many waypoints in images of Class 1, false detection still exists because of the scenes changing with time. Undetected is produced by the difference between training samples and testing samples. There is no undetected mistake in Class 2. It is because that there is no waypoint in the reference image of Class 2.

For comparison, the algorithms [2, 5] are used for waypoints selection. Random sampling investigation was carried out on sets of Class 1 and Class 2 with the same quantity of waypoints as in Table 1. The times are 1000 and the result is shown as in Table 4.

We can see that, in the analysis of Class 1, our method is better than the other two methods, and, in the analysis of Class 2, our method is better than [2] and almost the same as [5]. It is because that the threshold in [5] is set manually.

5. Conclusions

A method of waypoints selection was proposed in this paper, which firstly selected salient areas as candidate waypoints and then classified the candidates based on their feature attributes. The method combined the visual saliency analysis and feature attributes classification and especially avoided the inference of some small man-made areas with high brightness where there is no structure information.

The sensed image and the reference image are both from Google Earth, which makes the suitability analysis

TABLE 3: Classification rate.

Reference	Undetected	False detected
Class 1	15.78%	1.07%
Class 2	—	0.96%

TABLE 4: The comparison of results.

Result	Method		
	[2]	[5]	Proposed
Class 1			
Undetected	13.2%	22.25%	15.78%
False detected	10.46%	7.75%	1.07%
Class 2			
Undetected	—	—	—
False detected	2.03%	1.37%	1.43%

only depend on the original reference image itself and does not consider the matching condition under the geometrical transformations such as image scaling and rotation. In the next stage, we plan to extend this work along the following directions. Firstly, the matching condition will be considered for suitability analysis, and more aerial images under the real flying condition will be used to test the validity of the approach. Secondly, we will incorporate more powerful features or improve the saliency analysis model and classifying model to improve the effectiveness of the method.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

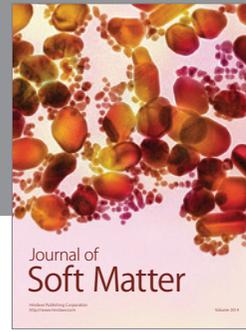
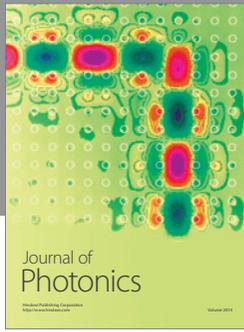
Acknowledgments

This research was financially supported by Xi’an Science and Technology Project (CXY1350(2)) and Aeronautical Science Foundation of China (20100853010).

References

- [1] L. Shen, Y. Bu, X. Xu, and L. Pan, “Research on matching-area suitability for scene matching aided navigation,” *Acta Aeronautica et Astronautica Sinica*, vol. 31, no. 3, pp. 553–563, 2010.
- [2] Z. Yang, Y. Chen, X. Qian, M. Yuan, and E. Gao, “Predicting the suitability for scene matching using SVM,” in *Proceedings of the International Conference on Audio, Language and Image Processing (ICALIP ’08)*, pp. 743–747, July 2008.
- [3] J. Du and T. Zhang, “Selection of matching region for scene matching,” *Infrared and Laser Engineering*, vol. 32, no. 4, pp. 368–371, 2003.
- [4] W.-X. Fu, J.-M. Wang, and S.-L. Jin, “Practical method for selecting scene matching area,” *Journal of Astronautics*, vol. 24, no. 4, pp. 348–353, 2003.

- [5] B. Jiang and Y. Chen, "Rule of selecting scene matching area," *Journal of Tongji University*, vol. 35, no. 6, pp. 830–833, 2007.
- [6] Y. Liu, F. Zhao, and S. Jin, "New method of selecting scene matching reference map," *Infrared and Laser Engineering*, vol. 30, no. 3, pp. 168–171, 2001.
- [7] R. An, X.-L. Jin, H.-L. Wang, X.-Z. Feng, and D.-X. Xu, "Image ability to obtain correct matching based on feature matching," *Infrared and Laser Engineering*, vol. 34, no. 4, pp. 469–473, 2005.
- [8] J. Han, P. Zhou, D. Zhang et al., "Efficient, simultaneous detection of multi-class geospatial targets based on visual saliency modeling and discriminative learning of sparse coding," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 89, pp. 37–48, 2014.
- [9] Y. Li, Y. Zhou, L. Xu, X. Yang, and J. Yang, "Incremental sparse saliency detection," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '09)*, pp. 3093–3096, November 2009.
- [10] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [11] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B. Methodological*, vol. 58, no. 1, pp. 267–288, 1996.
- [12] B. Han, H. Zhu, and Y. Ding, "Bottom-up saliency based on weighted sparse coding residual," in *19th ACM International Conference on Multimedia ACM Multimedia 2011, MM'11*, pp. 1117–1120, Scottsdale, Ariz, December 2011.
- [13] Z. C. Lin, M. M. Chen, L. Q. Wu, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," Tech. Rep. UILU-ENG-09-2215, University of Illinois at Urbana-Champaign, Champaign, Ill. USA, 2009.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

