KW-SIFT descriptor for remote-sensing image registration

Xiangzeng Liu (刘向增)*, Zheng Tian (田 铮), Weidong Yan (延伟东), and Xifa Duan (段西发)

School of Science, Northwestern Polytechnical University, Xi'an 710129, China

**Corresponding author: ccylxz20062008@126.com*

Received November 23, 2010; accepted January 14, 2011; posted online April 28, 2011

A technique to construct an affine invariant descriptor for remote-sensing image registration based on the scale invariant features transform (SIFT) in a kernel space is proposed. Affine invariant SIFT descriptor is first developed in an elliptical region determined by the Hessian matrix of the feature points. Thereafter, the descriptor is mapped to a feature space induced by a kernel, and a new descriptor is constructed by whitening the mapped descriptor in the feature space, with the transform called KW-SIFT. In a final step, the new descriptor is used to register remote-sensing images. Experimental results for remote-sensing image registration indicate that the proposed method improves the registration performance as compared with other related methods.

OCIS codes: 100.2000, 100.5010.
doi: 10.3788/COL201109.061001.

As a fundamental task in image processing, image registration is a process wherein two or more images of the same scene, taken at different times, from different viewpoints, or by different sensors are overlaid[1]. The existing image registration methods are, in general, divided into two broad categories: area-based and feature-based methods[2]. The feature-based methods do not work directly with image-intensity values but, instead, use salient features extracted from two images; this has been shown to be more suitable for situations when intensity changes and complicated geometric deformations are expected. Therefore, these feature-based methods have been used widely in remote-sensing image registration[3−5].

A feature-based application, syntactic descriptors for interest regions are successfully used in image registration because of their efficiency and robustness in geometrical and photometric transformations. Recent research work has focused on the development of descriptors invariant to image transformations because more distinctive and invariant descriptors would improve the performance of the feature-correspondence procedure[6,7]. Lowe developed a scale invariant features transform (SIFT) descriptor based on the gradient distribution in the detected regions; this descriptor is invariant to image scaling and rotation, and partially invariant to change in illumination[7]. As an extension of the SIFT method, Ke *et al.* presented a principal component analysis (PCA) SIFT descriptor for fast matching on the basis of the PCA technique[8], but this is less distinctive than the SIFT descriptor proved by Mikolajczyk *et al.*[9] More recently, Bay *et al.* proposed a novel scale-and-rotation invariant detector and descriptor, called the speeded up robust features (SURF), which was fast in matching and sufficiently distinctive[10]. To take advantage of the higher order statistical characteristics of image, Duan *et al.* applied an independent component analysis (ICA) on the gradient image around the detected interest point to develop ICA-SIFT descriptors[11], which are distinctive and have high matching speed. He *et al.* applied a new principle based on the system similarity theory to match SIFT descriptors and to obtain higher matching precision[12]. However, the performance of these algorithms drops rapidly as differences of input images grow in remote-sensing image registration. There are two reasons for that. Firstly, these descriptors are not invariant to affine transformation, which can induce the described regions to have certain different structures. Secondly, certain nonlinear correlations between descriptors from the same image are not considered, and this can affect the accuracy of matching.

In the original SIFT algorithm, the dominant gradient orientation is computed in a small circular neighborhood around the point. The size of the circular neighborhood is determined by the point's scale, but its shape is not invariant to affine transformation. Lindeberg *et al.* used the second moment matrix to estimate the elliptical region around the point, and proved that this fixed point would be preserved under affine transformations[13]. However, the algorithm is very complicated, whereas the Hessian matrix requires significantly less computational effort to solve than the second moment matrix; this is because the Hessian matrix only requires one scale parameter and is closely related to the determinant of the Hessian and the Laplacian functions[14]. In this letter, the Hessian matrix H is used to estimate the elliptical region around the point, which can be defined as

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}, \quad (1)$$

where D is the difference-of-Gaussian function:

$$D(x, y, \sigma) = G(x, y, k\sigma)^* I(x, y) - G(x, y, \sigma)^* I(x, y), \quad (2)$$

and $G(x, y, \sigma)$ is the Gaussian function: $G(x, y, \sigma) = \mathrm{e}^{-(x^2+y^2)/2\sigma^2}/2\pi\sigma^2$.

Theorem 1 Assume that there is a linear transformation $Y = AX$ between images I_1 and I_2; then, the elliptical region determined by the eigenvalues of the Hessian matrix for the corresponding feature points will be relatively invariant under affine transformations.

Proof Let H_1 and H_2 denote the Hessian matrix of the corresponding feature points, respectively. Similar to Ref. [13], $H_1 = A^{\mathrm{T}} H_2 A$ can be obtained. Let us assume that the eigenvalues of H_1,H_2, and A are α_1, α_2, β_1, β_2, and λ_1, λ_2, respectively. Then, the following equation

can be obtained:

$$\alpha_1\alpha_2 = \det(H_1) = \det(A^{\mathrm{T}}H_2A) = \det(AA^{\mathrm{T}}H_2) = \det(A^{\mathrm{T}}A)\cdot\det(H_2) = \lambda_1^2\lambda_2^2\beta_1\beta_2. \quad (3)$$

Therefore, the area of the elliptical affine regions determined by the eigenvalues of the two Hessian matrices can be computed as

$$S_1 = \pi\alpha_1\alpha_2, \quad S_2 = \pi\beta_1\beta_2, \quad S_1/S_2 = \lambda_1^2\lambda_2^2. \quad (4)$$

The ratio of the area of two elliptical affine regions is a constant that is related to the eigenvalues of the linear transformation matrix, which validates the result.

The eigenvalues of the Hessian matrix H represent two principal signal changes in a neighborhood of the point, and the elliptical region determined by the eigenvalues is relatively invariant under affine transformation according to Theorem 1. Therefore, these eigenvalues are used to measure the elliptical affine shape of the point neighborhood in our algorithm. To take advantage of the histogram that is used to determine the main orientation of feature points, the elliptical region is normalized to a circle by using the ellipse parameters from the Hessian matrix of the point. The position of each sample point that falls within the elliptical region can be mapped to its normalized position, within the circle, by the equation: $\bar{X} = H^{-1/2}X$.

Then, to construct a more robust descriptor, the Prewitt operator is used to compute the gradient magnitude $m(x,y)$ and gradient orientation $\theta(x,y)$ to avoid the use of the pixel differences directly, which can be computed as

$$P_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, \quad P_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad (5)$$

$$d_x(x,y) = P_x * L(x,y), \quad d_y(x,y) = P_y * L(x,y), \quad (6)$$

where $L(x,y,\delta)$ is produced from the convolution of a variable-scale Gaussian $G(x,y,\delta)$ with an input image $I(x,y)$: $L(x,y,\delta) = G(x,y,\delta) * I(x,y)$.

$$m(x,y) = \sqrt{(d_x(x,y))^2 + (d_y(x,y))^2}, \quad (7)$$

$$\theta(x,y) = \arctan(d_y(x,y)/d_x(x,y)). \quad (8)$$

After this, the remaining computations are similar to that in the SIFT. To reduce the computational complexity, the 4×4 descriptor with four orientations for a feature point is used.

However, certain nonlinear relationship may exist between the descriptors that form the same input image when the gray values or the geometrical structures of the areas around the two feature points are very similar; this results in a mismatch. The question then arises of how to eliminate the nonlinear relationship between the descriptors. The whitening PCA in the kernel space provides the answer. In our algorithm, the KW-SIFT descriptors can be obtained by using whitening of the ϕ-mapped descriptors in the feature space. Therefore, the KW-SIFT descriptors not only remove the correlations of the original descriptors but also reduce their dimensions.

Theorem 2 Let K be a kernel matrix with feature map ϕ and feature space F. For a given set of data $X=\{x_1,x_2,\cdots,x_n\}^{\mathrm{T}}\in\mathbb{R}^d$, the transpose of the data after whitening in the feature space is provided by the eigenvectors of the corresponding centered kernel matrix $\bar{K}$.

Proof Let $X_\phi = \{\phi(x_1),\phi(x_2),\cdots,\phi(x_n)\}^{\mathrm{T}}$ denote the images of X in the feature space and $K = X_\phi X_\phi^{\mathrm{T}}$ denote the corresponding kernel matrix with entries $K_{ij} = k(x_i,x_j) = \langle\phi(x_i),\phi(x_j)\rangle$. The data X_ϕ with its mean μ_ϕ can be shifted to obtain the centered data $\bar{X}_\phi = \{\bar{\phi}(x_1),\bar{\phi}(x_2),\cdots,\bar{\phi}(x_n)\}$ as

$$\bar{\phi}(x_i) = \phi(x_i) - \mu_\phi, \quad \mu_\phi = \frac{1}{n}\sum_{i=1}^{n}\phi(x_i). \quad (9)$$

Then, the covariance of $\bar{X}_\phi$ and the corresponding kernel matrix can be expressed as

$$C = \frac{1}{n}\sum_{i=1}^{n}\bar{\phi}(x_i)\bar{\phi}(x_i)^{\mathrm{T}} = \frac{1}{n}\bar{X}_\phi^{\mathrm{T}}\bar{X}_\phi, \quad \bar{K} = \bar{X}_\phi\bar{X}_\phi^{\mathrm{T}}. \quad (10)$$

The eigen-decomposition of nC and $\bar{K}$ are given by

$$nC = U\Lambda_C U^{\mathrm{T}}, \quad \bar{K} = V\Lambda_K V^{\mathrm{T}}, \quad (11)$$

where the columns u_i of the orthonormal matrix U are the eigenvectors of nC, and the columns v_i of the orthonormal matrix V are the eigenvectors of $\bar{K}$. Then, considering an eigenvector-eigenvalue pair v, λ of $\bar{K}$,

$$nC\bar{X}_\phi^{\mathrm{T}}v = \bar{X}_\phi^{\mathrm{T}}\bar{X}_\phi\bar{X}_\phi^{\mathrm{T}}v = \bar{X}_\phi^{\mathrm{T}}\bar{K}v = \lambda\bar{X}_\phi^{\mathrm{T}}v, \quad (12)$$

implying that $\bar{X}_\phi^{\mathrm{T}}v$, λ/n is an eigenvector-eigenvalue pair for C. Further, the norm of $\bar{X}_\phi^{\mathrm{T}}v$ is given by $\| \bar{X}_\phi^{\mathrm{T}}v \|^2 = v^{\mathrm{T}}\bar{X}_\phi\bar{X}_\phi^{\mathrm{T}}v = v^{\mathrm{T}}\bar{K}v = \lambda v^{\mathrm{T}}v = \lambda$.

Therefore, the normalized eigenvector of C is $u = \lambda^{-1/2}\bar{X}_\phi^{\mathrm{T}}v$. The k positive dominant eigenvalues of C are denoted by $\lambda_1/n \geqslant\cdots\geqslant \lambda_k/n > 0$ and the corresponding normalized eigenvectors by $u_i = \lambda^{-1/2}\bar{X}_\phi^{\mathrm{T}}v_i, i = 1,\cdots,k$. Let $D = \mathrm{diag}(\lambda_1/n,\cdots,\lambda_k/n)$ and $E = (u_1,\cdots,u_k)$, then the whitening matrix can be calculated via

$$\begin{aligned} B &= D^{-1/2}E^{\mathrm{T}} \\ &= \{(\lambda_1/n)^{-1/2}\lambda_1^{-1/2}v_1^{\mathrm{T}}\bar{X}_\phi,\cdots,(\lambda_k/n)^{-1/2}\lambda_k^{-1/2}v_k^{\mathrm{T}}\bar{X}_\phi\}^{\mathrm{T}} \\ &= n^{1/2}\{\lambda_1^{-1}\sum_{i=1}^{n}v_1^i\bar{\phi}(x_i),\cdots,\lambda_k^{-1}\sum_{i=1}^{n}v_k^i\bar{\phi}(x_i)\}^{\mathrm{T}}. \end{aligned} \quad (13)$$

Therefore, the jth column of the whitened matrix Z^{T} is

$$\begin{aligned} Z_j^T = B\bar{X}_\phi^j &= n^{1/2}\{\lambda_1^{-1}\sum_{i=1}^{n}v_1^i\bar{\phi}^{\mathrm{T}}(x_i)\bar{\phi}(x_j), \\ &\quad \cdots,\lambda_k^{-1}\sum_{i=1}^{n}v_k^i\bar{\phi}^{\mathrm{T}}(x_i)\bar{\phi}(x_j)\}^{\mathrm{T}} \\ &= n^{1/2}\{\lambda_1^{-1}\sum_{i=1}^{n}v_1^i\bar{k}(x_i,x_j),\cdots,\lambda_k^{-1}\sum_{i=1}^{n}v_k^i\bar{k}(x_i,x_j)\}^{\mathrm{T}} \\ &= n^{1/2}\{v_1^j,\cdots,v_k^j\}^{\mathrm{T}}, \quad j = 1,2,\cdots,n, \end{aligned} \quad (14)$$

where

$$\bar{k}(x,z) = k(x,z) - \frac{1}{n}\sum_{i=1}^{n}k(x,x_i) - \frac{1}{n}\sum_{i=1}^{n}k(z,x_i) + \frac{1}{n^2}\sum_{i,j=1}^{n}k(x_i,x_j). \quad (15)$$

Then, Z can be depicted as $Z = n^{1/2}\{v_1, \cdots, v_k\}$, where $v_1, \cdots, v_k$ are the corresponding eigenvectors of the first k positive dominant eigenvalues of $\bar{K}$.

The construction of the KW-SIFT descriptors can be divided into two steps. Firstly, the affine invariant descriptors are mapped to a feature space via certain nonlinear mapping ϕ; then, the kernel matrix K can be calculated. Secondly, the centered kernel matrix $\bar{K}$ is calculated and its first k eigenvectors can be found, which are the KW-SIFT descriptors in accordance with Theorem 2. In our algorithm, the Gaussian kernel of the form $k(x, y) = \exp(- \parallel x - y \parallel^2 / 2\sigma^2)$ is employed.

The proposed algorithm for remote-sensing image registration by using the KW-SIFT descriptor can be expressed as follows.

1) Construct the affine invariant descriptors $X = \{x_1, x_2, \cdots, x_n\}^{\mathrm{T}}$ for the extracted feature points from the input images I_1 and I_2.

2) Map those descriptors to a feature space and calculate the corresponding KW-SIFT descriptors $Z = n^{1/2}\{v_1, v_2, \cdots, v_k\}$.

3) The Euclidian distance between new descriptors, which consist of kernel principal components of original ones, is employed to determine the corresponding feature-point pairs.

4) Compute the transformation parameters between the input images I_1 and I_2 according to the corresponding feature-point pairs, which are used to complete the image registration.

In the proposed algorithm described above, the dimension of the affine invariant descriptor is $n = 64$. In the experiments, the dimension of the KW-SIFT descriptor is taken as $k = 32$ and the ratio of scales between the two feature points is used as the value of parameter σ in the Gaussian kernel matrix. The transformation between the input images is assumed to be an affine transformation, which can be defined as

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} a & b \\ d & e \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c \\ f \end{bmatrix}, \tag{16}$$

where (x, y) is the point of the reference image, (X, Y) is the point of the sensed image, and a, b, c, d, e, f are the transformation parameters.

To verify the validity of the proposed algorithm, the experimental results are divided into two parts. In the first part, the algorithm is applied to register two synthetic aperture radar (SAR) images (Fig. 1) whose transformation parameters are known to evaluate the accuracy of the registration approach. In the second part, the proposed algorithm is used to register three different pairs of remote-sensing images (Fig. 2) and compare it with other related methods. The root mean-square error (RMSE) between the matched feature points is used for evaluating the registration results, which is defined as

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{m} [(ax_i + by_i + c - X_i)^2 + (dx_i + ey_i + f - Y_i)^2] / m}, \tag{17}$$

where m represents the total number of matched feature-point pairs. In this study, the correspondence of feature points was obtained by comparing the distance of the closest neighbor descriptor to that of the next closest

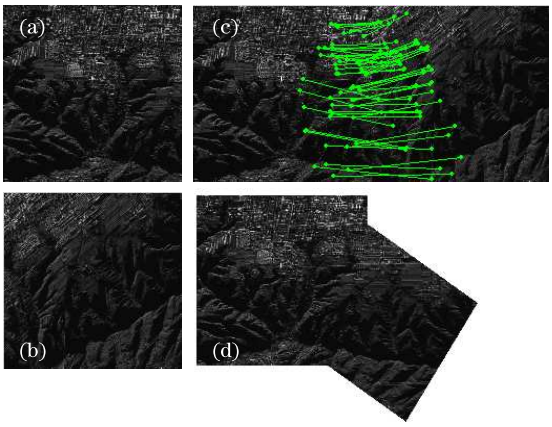

Fig. 1. Registration of SAR images from some region of Yunnan in China. (a) Reference image; (b) sensed image; (c) matching results; (d) registration results.

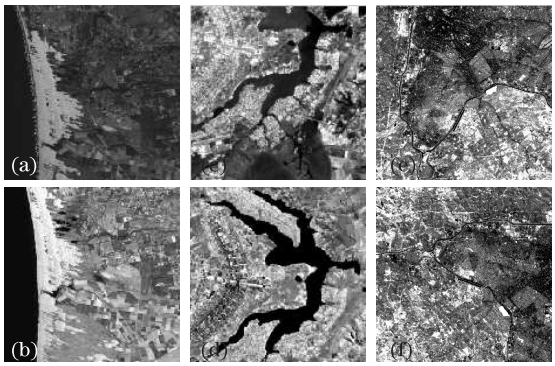

Fig. 2. Three pairs of remote-sensing images with large variations. (a), (c), and (e) are the reference images. (b), (d) and (f) are the corresponding sensed images.

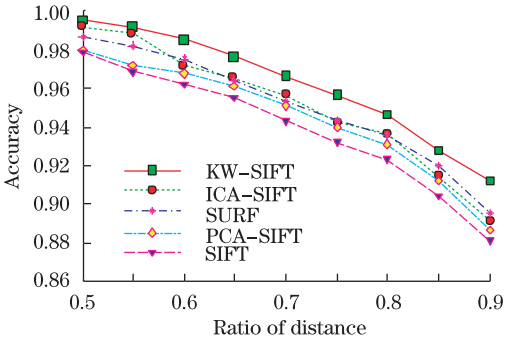

Fig. 3. Comparison of matching accuracy of Figs. 1(a) and (b) by five related methods when the ratio of distance changes from 0.5 to 0.9.

descriptor. To develop a compromise between the number of correct matched points and the correct match rate, the threshold was set to 0.6. The threshold $r = 0.6$ is considered and the two feature points are treated as matched pairs if the ratio of distance is less than the threshold. A pair of feature points is considered to be the correct matched points if their matching error is less than 0.5. In the experiments, 20 matched pairs were used to calculate the RMSE for different methods.

In order to demonstrate that the proposed algorithm is robust with regard to large geometric variations, experiments have been carried out on two SAR images in comparing the proposed algorithm (KW-SIFT) with SIFT[7],

PCA-SIFT[8], SURF[10], and ICA-SIFT[11]. The two SAR images (Figs. 1(a) and (b)) with large geometric variations are obtained from some regions of the Yunnan Province in China, and they have been used to test the accuracy of the proposed algorithm. Figures 1(a) and (b) are used as the reference and the sensed images, respectively. Figure 1(b) represents the affine transformed version of Fig. 1(a), where the transformation parameters are known. The matching results and the registration results obtained by using this new algorithm are shown in Figs. 1(c) and (d), respectively. From the results, it can be seen that the two images have a large affine transformation. The comparison of matching accuracy for different methods when the ratio of distances changes from 0.5 to 0.9 is shown in Fig. 3. Here, the matching accuracy is defined as the ratio between the number of correct matches and the total number of matches detected. The results indicate that the matching accuracy of our methods is higher than 90% and decreases more slowly with the increase in ratio of distances, in comparison with other related methods. The actual transformation parameters and those obtained by the related methods are shown in Table 1. From Table 1, it is clear that differences between the actual transformation parameters and that of the proposed algorithm are the least. According to the actual transformation parameters, the RMSEs of the registration results by different methods are also given in Table 1. Clearly, the image registration by using the proposed algorithm results in a lower RMSE, and its accuracy, therefore, is better than that of the other four related methods.

To test the registration ability of the proposed algorithm, three remote-sensing image pairs with different changes were used. The first pair consists of images from the same sensor (Landsat TM) but acquired from different bands, as shown in Figs. 2(a) (Band 3) and (b) (Band 5), which have large intensity variations. The images of the second and the third pairs from different sensors acquired at different times (Figs. 2(c)–(f)), in which changes can be seen to occur between Figs. 2(c) and (d), and Figs. 2(e) and (f), have scaling and rotational variations. Figures 2(c) and (e) are spot (Band 3) images obtained on Aug. 8, 1995, whereas Figs. 2(d) and (f) are Landsat TM (Band 4) images obtained on Jun. 7, 1994. The registration results obtained by using the proposed algorithm for the three image pairs are shown in Figs. 4(a)–(c). Comparisons of matching accuracy for the three image pairs when the ratio of distances changes from 0.5 to 0.9 are shown in Figs. 5(a)–(c). From Fig. 5, it can be seen that the matching accuracies for the three image pairs by using the proposed algorithm are higher than 75% and decrease more slowly than the other four related methods with changes in the ratio of distances. Although the proposed algorithm provides slightly better results than those obtained with the other four related methods when structures of the image pairs are not very complex (Fig. 5(a)), it has obvious advantages in comparison with the others when the complexity of structures increases (Figs. 5(b) and (c)); this indicates that the proposed algorithm is more robust than other related methods when image pairs have complex structures

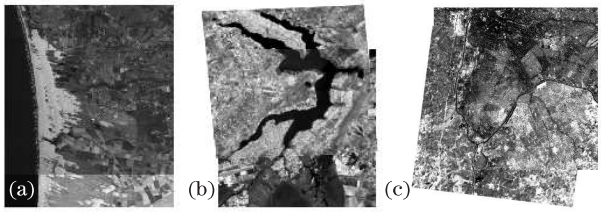

Fig. 4. Registration results for the three image pairs in Fig. 2 by using the proposed algorithm. (a) Registration results for Figs. 2(a) and (b); (b) registration results for Figs. 2(c) and (d); (c) registration results for Figs. 2(e) and (f).

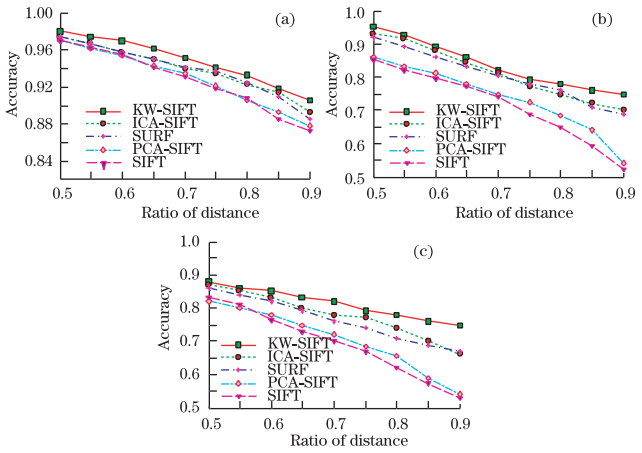

Fig. 5. Comparison of matching accuracy for the three image pairs in Fig. 2 by five related methods when the ratio changed from 0.5 to 0.9. (a) Matching accuracy of Figs. 2(a) and (b); (b) matching accuracy of Figs. 2(c) and (d); (c) matching accuracy of Figs. 2(e) and (f).

and large variations. The RMSEs of the registration results for three image pairs by using different methods are shown in Table 2, indicating that the proposed algorithm outperforms the other four related methods.

The performance of the proposed algorithm is analyzed in terms of the complexity and the application scope for affine transformation. Estimating the complexity of the proposed algorithm mainly boils down to calculation of

Table 1. Comparison of Registration Results Obtained by Different Methods

Methods	a	b	c	d	e	f	RMSE
Actual	0.8300	0.5000	–348.7500	–0.7200	1.0000	283.9700	0.0000
SIFT	0.8219	0.4804	–350.3315	–0.7177	0.9892	284.6600	0.8879
PCA-SIFT	0.8315	0.5152	–349.8800	–0.7305	1.0332	283.8256	0.7126
SURF	0.8308	0.4905	–348.0228	–0.7348	1.0556	283.7306	0.2879
ICA-SIFT	0.8278	0.4822	–349.2537	–0.7116	0.9961	285.5532	0.2836
KW-SIFT	0.8293	0.4998	–348.7252	–0.7201	1.0001	283.9600	0.0512

Table 2. RMSEs of Registration Results Using the Five Different Methods for Fig. 2

Method	SIFT	PCA-SIFT	SURF	ICA-SIFT	KW-SIFT
1st Pair	0.8562	0.7973	0.5774	0.4523	0.2526
2nd Pair	1.2538	1.2506	0.8611	0.6461	0.4833
3rd Pair	1.2675	1.2466	0.9539	0.7877	0.4957

Table 3. Run Time and Number of Correct Matching Pairs Using Different Methods for Fig. 2

Method		SIFT	PCA -SIFT	ICA -SIFT	SURF	KW -SIFT
Run Time (s)	1st Pair	5.2	4.3	5.6	5.0	5.5
	2nd Pair	3.1	2.8	3.5	2.9	3.0
	3rd Pair	6.5	5.7	6.2	6.1	6.3
Number of Correct Match Pairs	1st Pair	51	55	60	57	72
	2nd Pair	6	10	15	18	20
	3rd Pair	22	21	31	32	38

the KW-SIFT descriptors and matching of the descriptors. Let us assume that the sizes of the two input images are M and N ($M \leqslant N$), and the numbers of feature points extracted from the two images are m and n ($m \leqslant n$), respectively. Then, the complexities involved in the calculation of the affine invariant SIFT and the KW-SIFT descriptors are $O(N^2)$ and $O(n^3)$, respectively, and that of matching the descriptors is $O(mn)$. Indeed, the complexity of the proposed method is proportional to the input image area and the number of extracted feature points. Table 3 presents a comparison of the computation time required by the different methods. Here, each method is applied to the three pairs of remote-sensing images (Fig. 2). The experiments were performed on a standard computer (Pentium E5400, running at 2.70 GHz). The proposed algorithm obtains the most correct matches although its registration speed is slightly lower than those for other methods; therefore, it achieves a compromise between speed and quality of image registration. To test the conditions where the proposed algorithm can be applied, it was applied for the registration of Fig. 1(a) and the different affine transform versions of that particular image. The results indicate that the proposed algorithm can register the two images whose scales ($s = \sqrt{a^2 + b^2}$) change within 3.5 times, whereas the SIFT can only deal with scale variation of less than 2 times. The rotation ($\theta = \arccos(a/s)$) and translation (e, f) parameters have little effect on the results.

In conclusion, we present a technique to construct an affine invariant descriptor for remote-sensing image registration based on the SIFT in kernel space, namely, KW-SIFT. An affine invariant SIFT descriptor is developed based on the SIFT descriptor. The new descriptor KW-SIFT descriptor is constructed by whitening the mapped descriptor in the kernel space. According to Theorem 2, the descriptors are the components of the eigenvectors of the kernel matrix, which remove the correlation of the original descriptors and improve the accuracy of the registration results. The experimental results on remote-sensing image pairs with different variations indicate that the proposed method returns better performance as compared with SIFT, PCA-SIFT, SURF, and ICA-SIFT methods with regard to matching accuracy and RMSE of the registration results.

This work was supported by the National Natural Science Foundation of China (Nos. 60972150 and 10926197).

References

1. L. G. Brown, ACM Comput. Surv. **24,** 325 (1992).
2. B. Zitová and J. Flusser, Image Vis. Comput. **21,** 977 (2003).
3. Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, IEEE Trans. Geosci. Remote Sens. **43,** 2127 (2005).
4. Z. Chen, F. Yin, and F. Sun, Acta Opt. Sin. (in Chinese) **29,** 2744 (2009).
5. C. Leng, Z. Tian, J. Li, and M. Ding, Chin. Opt. Lett. **7,** 996 (2009).
6. Q. Zeng, L. Liu, and J. Li, Chin. Opt. Lett. **8,** 573 (2010).
7. D. G. Lowe, Int. J. Comput. Vis. **60,** 91 (2004).
8. Y. Ke and R. Sukthankar, in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 511 (2004).
9. K. Mikolajczyk and C. Schmid, IEEE Trans. Pattern Anal. Mach. Intell. **27,** 1615 (2005).
10. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, Comput. Vis. Image Underst. **110,** 346 (2008).
11. C. Duan, X. Meng, C. Tu, and C. Yang, IET Comput. Vis. **2,** 178 (2008).
12. J. He, J. Yang, B. Xue, X. Chen, and P. Yan, Acta Opt. Sin. (in Chinese) **30,** 989 (2010).
13. T. Lindeberg and J. Gårding, Image Vis. Comput. **15,** 415 (1997).
14. R. Lakemond, C. B. Fookes, and S. Sridharan, in *Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance* 496 (2004).