Spectral feature matching based on partial least squares

Weidong Yan (延伟东)[1*], Zheng Tian (田 铮)[1,2], Lulu Pan (潘璐璐)[1], and Mingtao Ding (丁明涛)[1]

[1] *School of Science, Northwestern Polytechnical University, Xi'an 710072*

[2] *State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101*

[*] *E-mail: weidongyan@qq.com*

Received September 16, 2008

We investigate the spectral approaches to the problem of point pattern matching, and present a spectral feature descriptors based on partial least square (PLS). Given keypoints of two images, we define the position similarity matrices respectively, and extract the spectral features from the matrices by PLS, which indicate geometric distribution and inner relationships of the keypoints. Then the keypoints matching is done by bipartite graph matching. The experiments on both synthetic and real-world data corroborate the robustness and invariance of the algorithm.

OCIS codes: 100.0100, 100.2000.

doi: 10.3788/COL20090703.0201.

Image registration is a fundamental task in image processing used to match two or more images which are taken at different time from different sensors or different viewpoints, so that differences can be detected. The most difficult part of a registration process is the determination of the correspondence of keypoints between the images to be registered. If some correspondences are incorrect, they will produce an incorrect transformation function, which could yield totally wrong results. So a highly robust point matching algorithm is needed.

Image registration methods can be generally organized in area-based and feature-based methods[1]. Area-based methods deal with the images without attempting to detect salient features, and adopt optimization algorithms[2,3]. These methods have some limitations, which can be affected by the intensity distribution. Feature-based matching which is particularly suited to multi-source image registration consists of three stages. In the first stage, features in the image are extracted, such as keypoints, lines, and patches. In the second stage, keypoints in the reference image are corresponded with keypoints in the sensed image. In the last stage, a spatial mapping, usually an affine transformation, is determined using these matched keypoints based on least square regression or similar techniques. However, in most cases, the extraction and representation of the relationship itself are difficult problems. The crucial objective[4] of all feature-based matching methods is to have discriminative and robust feature descriptors that are invariant to all assumed differences between the images.

Keypoints are the simplest form of features, which are represented basically by the point locations. However, the resulting point matching problem can be quite difficult because of various factors like noise. Keypoints can be matched by considering either the radiometric properties of the surrounding pixels, or the geometric distribution of the whole set of keypoints across the whole image[5]. Spectral graph theory is a term applied to a family of techniques that aim to characterize the global structural properties of graphs using the eigenvalues and eigenvectors of similarity matrices[6]. So the spectral graph can indicate geometric distribution of keypoints. In the computer vision literatures, there have been a number of attempts to use spectral properties for graph-matching[7−9]. Scott *et al.* used a Gaussian weighting function to build an inter-image similarity matrix between feature points in different images being matched and then performed singular value decomposition (SVD) on the similarity matrix in order to get correspondences from the similarity matrix's singular values and vectors[7]. This method fails when the rotation or scaling between the images is too large. To overcome this problem, Shapiro *et al.*[8] constructed intra-image similarity matrices for the individual points-sets being matched with an aim to capture the relational image structure[8]. The eigenvectors of the individual similarity matrices were used to match. This method can be viewed as projecting the individual point-sets into an eigenspace, and seeking matches by looking for the closest point correspondence. Wang *et al.* investigated the performance of kernel principal component analysis (PCA) with a polynomial kernel function for solving the point correspondence problem and discussed the relationship with Shapiro's correspondence method[9]. Such approaches characterize the graphs by their dominant eigenvectors. However, these eigenvectors are computed independently for each graph and thus often do not capture co-salient structures of the graphs. While the partial least square (PLS) approach helps to extract representations from two images which contain relevant information for the matching of the particular pair of images.

In this letter, we present a method of spectral feature matching based on PLS. The spectral features are constructed with information of position similarity matrices, using PLS components, and invariant to translation, scale, and rotation, and very suitable for feature-based matching.

The PLS method initially developed by Wold *et al.*[10] has a tremendous success in chemometrics and chemical industries for static data analysis. It integrates the PCA and canonical correlation analysis (CCA) together naturally and is convenient for the analysis of the multi-

dimensional complexity system. In its general form, PLS creates components by using the existing correlations between different sets of variance while also keeping most of the variance of both sets. Before detailing the algorithm, we provide some of the formal ingredients of the method.

Consider a general setting of the PLS algorithm to model the relation between two data sets. Let $x = (x_1, \cdots, x_N)$ denote an N-dimensional vector of variables in the first block of data, and similarly let $y = (y_1, \cdots, y_M)$ denote a vector of variables from the second set. Observing n data samples from each block of variables, PLS decomposes $X = (x_{ij})_{n\times N}$ and $Y = (y_{ij})_{n\times N}$ into the form

$$X = TP^{\mathrm{T}} + F,$$

$$Y = UQ^{\mathrm{T}} + G,$$

where T and U are $n \times r$ matrices of the extracted r components, the $N \times r$ matrices P and Q represent matrices of projections, and the $n \times N$ matrices F and Q are the matrices of residuals. The PLS method, in which the classical form is based on the nonlinear iterative partial least square (NIPALS) algorithm, finds projection axes w and c such that[10]

$$\max \quad S = t^{\mathrm{T}}u_{=}(Xw)^{\mathrm{T}}(Yc) = w^{\mathrm{T}}X^{\mathrm{T}}Yc,$$

$$\text{s.t.} \quad \begin{cases} w^{\mathrm{T}}w = \|w\|^2 = 1 \\ c^{\mathrm{T}}c = \|c\|^2 = 1 \end{cases}. \tag{1}$$

The solution to this optimization problem is given by the following eigenvalue problem[11]:

$$X^{\mathrm{T}}YY^{\mathrm{T}}Xw = \lambda w,$$

where λ is the eigenvalue associated with w. The components of X are then given as $t = Xw$.

Similarly, the extraction of components of Y is given as

$$XX^{\mathrm{T}}YY^{\mathrm{T}}t = \lambda t, \tag{2}$$

$$u = YY^{\mathrm{T}}t. \tag{3}$$

Now, we select the keypoint sets on the reference image and the sensed image respectively, and denote $X = (x_1, x_2, \cdots, x_n)$ and $Y = (y_1, y_2, \cdots, y_n)$. In this letter, the keypoints in each data-set are in the form of $x_i = (x_i^1, x_i^2)$ and $y_i = (y_i^1, y_i^2)$, respectively. Our aim is to establish a one-to-one point correspondence between the two data-sets.

Using the keypoint sets, we construct the position similarity matrices by Gaussian kernel function, $(S_x)_{n\times n}$ and $(S_y)_{n\times n}$,

$$(S_x)_{ij} = \exp(-\frac{d(x_i, x_j)^2}{\sigma_x^2}),$$

$$\text{and} \qquad (S_y)_{ij} = \exp(-\frac{d(y_i, y_j)^2}{\sigma_y^2}),$$

where $d(x_i, x_j)$ is the Euclidean distance between the keypoints x_i and x_j, and σ_x, σ_y are adjustable parameters. Now every keypoint corresponds to an n-dimensional feature vector, $x_i \to (S_x)_{i\cdot}$, $y_j \to (S_y)_{j\cdot}$. The mapping of the original two-dimensional (2D) data to a higher dimensional space is completed and thus the structural information can be captured from the feature vectors[8].

Under the criterion (1), the number of components is $\mathrm{rank}(S_x^{\mathrm{T}}S_yS_y^{\mathrm{T}}S_x)$ (that is the number of non-zero eigenvalues of matrix $S_x^{\mathrm{T}}S_yS_y^{\mathrm{T}}S_x$) pairs at most. r ($\leq$ $\mathrm{rank}(S_x^{\mathrm{T}}S_yS_y^{\mathrm{T}}S_x)$) pairs of components are composed of vectors which are selected from the eigenvectors corresponding to the first r maximum eigenvalues of eigenequations

$$S_xS_x^{\mathrm{T}}S_yS_y^{\mathrm{T}}T = \lambda^2 T, \tag{4}$$

$$S_yS_y^{\mathrm{T}}S_xS_x^{\mathrm{T}}U = \lambda^2 U, \tag{5}$$

where $T = [t_1, \cdots, t_r]$, $U = [u_1, \cdots, u_r]$.

Now we give the proof of the above theorem. Using the Lagrange multiplier method[12] to transform Eq. (1), we get

$$L(W, C) = W^{\mathrm{T}}S_x^{\mathrm{T}}S_yC - \frac{\lambda_1}{2}(W^{\mathrm{T}}W - 1) - \frac{\lambda_2}{2}(C^{\mathrm{T}}C - 1),$$

where λ_1 and λ_2 are Lagrange multipliers. Let

$$\frac{\partial L(W, C)}{\partial W} = S_x^{\mathrm{T}}S_yC - \lambda_1 W = 0,$$

$$\frac{\partial L(W, C)}{\partial C} = S_y^{\mathrm{T}}S_xW - \lambda_2 C = 0,$$

then,

$$W^{\mathrm{T}}S_x^{\mathrm{T}}S_yC = \lambda_1 W^{\mathrm{T}}W = \lambda_1,$$

$$C^{\mathrm{T}}S_y^{\mathrm{T}}S_xW = \lambda_2 C^{\mathrm{T}}C = \lambda_2.$$

Since $(S_y^{\mathrm{T}}S_x)^{\mathrm{T}} = S_x^{\mathrm{T}}S_y$, we have

$$\lambda_1 = \lambda_1^{\mathrm{T}} = (W^{\mathrm{T}}S_x^{\mathrm{T}}S_yC)^{\mathrm{T}} = C^{\mathrm{T}}S_y^{\mathrm{T}}S_xW = \lambda_2.$$

Let $\lambda_1 = \lambda_2 = \lambda$, we can infer that to obtain the maximum value of λ is the same as to maximize the criterion (1).

Then r pairs of projection axes W, C can be inferred via

$$S_x^{\mathrm{T}}S_yS_y^{\mathrm{T}}S_xW = \lambda^2 W,$$

$$S_y^{\mathrm{T}}S_xS_x^{\mathrm{T}}S_yC = \lambda^2 C.$$

Multiplying both side of Eq. (4) by S_x, then the components of S_x are given by $T = S_xW$, that is $S_xS_x^{\mathrm{T}}S_yS_y^{\mathrm{T}}T = \lambda^2 T$. Since both $S_x^{\mathrm{T}}S_yS_y^{\mathrm{T}}S_x$ and $S_y^{\mathrm{T}}S_xS_x^{\mathrm{T}}S_y$ are symmetric matrices, and $\mathrm{rank}(S_x^{\mathrm{T}}S_yS_y^{\mathrm{T}}S_x) = \mathrm{rank}(S_y^{\mathrm{T}}S_xS_x^{\mathrm{T}}S_y)$, the two eigenequations (4) and (5) have the same nonzero eigenvalues.

In the same way, $S_yS_y^{\mathrm{T}}S_xS_x^{\mathrm{T}}U = \lambda^2 U$.

Now the keypoint x_i in the reference image and y_j in the sensed image can be represented by PLS spectral features $(T)_i = [t_{i1}, t_{i2}, \cdots, t_{ir}]$ and $(U)_j =$

$[u_{j1}, u_{j2}, \cdots, u_{jr}]$ in the r-dimensional eigenspace spanned by T and U respectively. Thus if the arbitrary numbering of two features in an image is changed, their feature vectors simply change positions in T (or U), and the matching of keypoints can be converted to the matching of PLS spectral features. It is known from geometry that the Euclidean distance is invariant to the similarity transformation[12]. Hence, the spectral features have invariance.

Various methods have been proposed to deal with the keypoint matching problem[13,14]. As a fundamental problem in image registration, graph matching has a variety of applications in the field of computer vision. In graph matching, keypoints are modeled as graphs and feature matching amounts to find a correspondence between the nodes of different graphs. In this letter, bipartite matching methods are applied in feature matching. We firstly extract keypoints from the reference and sensed images, and then compute the spectral feature vectors (T and U) for the keypoints, forming the vertices for a bipartite graph. Each vertex is connected to all the vertices on the opposite side. The similarity measures between these two sets of spectral features are summarized in the association matrix. We then apply the Hungarian algorithm to find the optimal matching for the bipartite graph.

To make the algorithm more robust, the mismatching keypoints can be eliminated by the continuity constraint (close neighbors in the reference image must be mapped to close neighbors in the sensed image). So we can examine the correspondences of the neighbors of the matched keypoints to eliminate the mismatching.

We investigate the performance of the method of spectral feature matching based on PLS. We use it to solve a 2D rigid mapping (rotation, translation, and scale). Three experiments were performed. In the experiment 1, we compared the matching performance of the PLS-based approach described above and SVD-based approach[7], and showed the robustness of the PLS-based approach. In the experiment 2, we verified the invariance of the spectral features on synthetic and optical images. Finally, in the experiment 3, we tested the method on a synthetic aperture radar (SAR) image.

Firstly, we investigated the effect of controlled affine skew of the point sets. The reference point set was randomly generated and then transformed by parameters to get the sensed data set. Figure 1 shows the matching results of SVD-based matching and PLS-based feature matching. Tables 1 and 2 show the accuracy results of PLS-based feature matching and SVD-based matching for four different transform models. In Fig. 1 and the tables, we define the known translational transformations in x-axis and y-axis as dx and dy, the rotational transformation as θ, and the scale transformation as s. From the results, it is obvious that our technique permits highly accurate matching results, especially when the parameter θ is large. The SVD-based algorithm does not cope with large rotation in the image. Small root-mean-square (RMS) errors have been observed in most cases of different transform models.

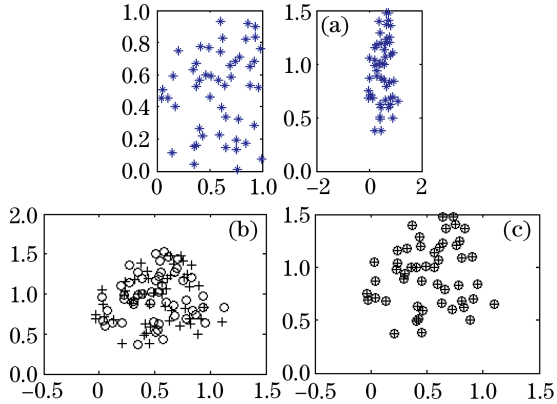

Fig. 1. (a) Reference data set and sensed data set without outliers; (b) SVD-based matching result; (c) PLS-based feature matching result. In (b) and (c), the circles are the reference data, the crosses are the sensed data. dx = 0.1, dy = 0.2, $\theta = 20^\circ$, $s = 1.1$.

Table 1. Accuracy of PLS-Based Feature Matching Using 50 Point-Sets

Ground Truth					Matching Results			
dx	dy	θ (rad)	s	Data Sets	dx	dy	θ (rad)	s
0.4	0.5	0.1745	1.1	50	0.4	0.5	0.1745	1.1
0.1	0.2	0.1745	1.2	50	0.1	0.2	0.1745	1.2
0.1	0.2	0.2618	1.1	50	0.1	0.2	0.2618	1.1
0.1	0.2	0.5236	1.1	50	0.1	0.2	0.5236	1.1
RMS Error					0	0	0	0

Table 2. Accuracy of SVD-Based Feature Matching Using 50 Point-Sets

Ground Truth					Matching Results			
dx	dy	θ (rad)	s	Data Sets	dx	dy	θ (rad)	s
0.4	0.5	0.1745	1.1	50	0.4003	0.4999	0.1746	1.0992
0.1	0.2	0.1745	1.2	50	0.1035	0.2111	0.1680	1.1930
0.1	0.2	0.2618	1.1	50	0.1237	0.2128	0.2824	1.0845
0.1	0.2	0.5236	1.1	50	0.4307	0.2698	0.9606	1.0567
RMS Error					0.1658	0.0361	0.2188	0.0233

Secondly, we focused on the performance of the algorithms when the data were under affine transformations and contained uncertainties such as outliers and noise. For this purpose, we added noise to the sensed data. The reference data set was the same as that in the experiment 1, and 5 outliers were added to the sensed data set. The results are shown in Fig. 2. The PLS-based algorithm presents a strong ability to eliminate incorrect matching. And the total computation time elapsed during the matching process is less than 1 s for 50 points matching.

To provide more quantitative evaluations, we also tested the algorithm on synthetic and optical images. We matched images from a gesture of a hand. The keypoints in hand images are points of maximum curvature on the outline of the hand. Figure 3 shows the final configuration of correspondence matches obtained using our method. We also matched optical images from a house. There are rotation and scaling distortions. Figure 4 shows the final configuration of correspondence matches obtained using our method. From these results,

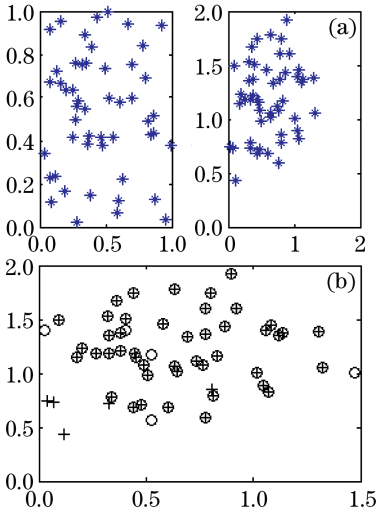

Fig. 2. (a) Reference data set and sensed data set with outliers; (b) the matching results by our implementation found between the two data sets, the circles are the reference data, the crosses are the sensed data.

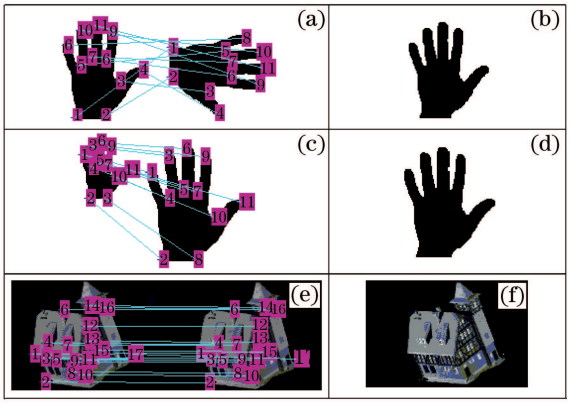

Fig. 3. Point matching results on the synthetic images. (a) Correspondences between the hand-roatation and (b) registration result; (c) correspondences between the hand-scaling and (d) registration result; (e) correspondences between the houses and (f) registration result.

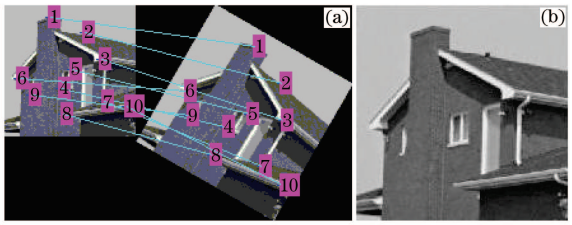

Fig. 4. Synthetic experiment (rotation). (a) Correspondences of keypoints between the images and (b) registration result.

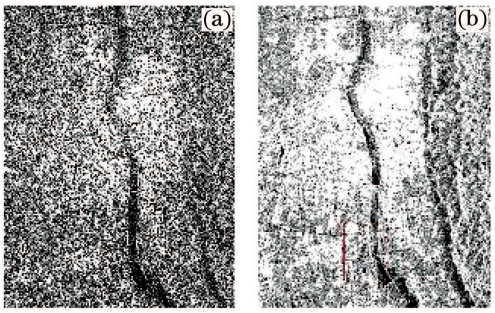

Fig. 5. Deyang area. (a) Unregistered image before the earthquake (Dec. 19, 2007); (b) unregistered image after the earthquake (May 14, 2008).

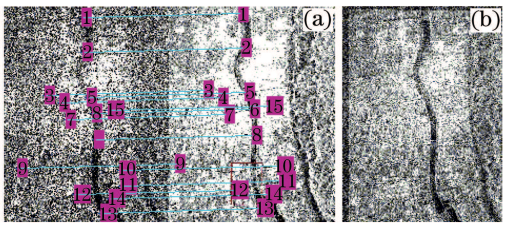

Fig. 6. SAR experiment. (a) Correspondences of keypoints between the images; (b) registration result computed with the proposed approach.

we can see that the spectral descriptors are invariant to rotation and scaling.

To test our algorithm on a real-world situation, we applied it to a SAR image registration problem. We took two images from the same area (Deyang) with 15 keypoints already extracted. Then we tried to match using our implementation. Figure 5 shows two SAR images taken from the city of Deyang by RADARSAT-1 satellite, before (Dec. 19, 2007) and after (May 14, 2008) the devastating earthquake of May 12, 2008. Figure 6 shows the registration result.

In conclusion, we investigate the spectral approaches to the problem of point pattern matching. Firstly, we have considered the rigid point-set alignment. PLS can be effectively used for solving the rigid point correspondence matching problem. The PLS components from both the reference and sensed images are used as spectral descriptors to establish their inner relationships. The spectral descriptors indicate geometric distribution of keypoints and are invariant to translation, scale, and rotation. Secondly, we have used robust methods for point correspondences by the continuity constraint. Further work is needed to obtain more robust results using more matching keypoints to get better mapping approximations and work with keypoint-sets of different sizes.

This work was supported by the Northwestern Polytechnical University Doctoral Dissertation Innovation Foundation (No. CX200819), the National Natural Science Foundation of China (No. 60375003), the Astronautics Basal Science Foundation of China (No. 03I53059), and the Science and Technology Innovation Foundation of the Northwestern Polytechnical University (No. 2007KJ01033).

References

1. B. Zitová and J. Flusser, Image Vision Comput. **21,** 977 (2003).
2. X. Yang and J. Pei, Chin. Opt. Lett. **3,** 510 (2005).
3. F. Ye, L. Su, and S. Li, Chin. Opt. Lett. **4,** 386 (2006).
4. W. Aguilar, Y. Frauel, F. Escolano, M. E. Martinez-Perez, A. Espinosa-Romero, and M. A. Lozano, Image Vision Comput. (to be published).
5. P. Dare and I. Dowman, ISPRS J. Photogramm. Remote Sens. **56,** 13 (2001).
6. M. Carcassoni and E. R. Hancock, Pattern Recogn. **36,** 193 (2003).
7. G. L. Scott and H. C. Longuet-Higgins, Proc. R. Soc. Lond. B **244,** 21 (1991).
8. L. S. Shapiro and J. M. Brady, Image Vision Comput. **10,** 283 (1992).
9. H. Wang and E. Hancock, Lecture Notes in Computer Science **3138,** 361 (2004).
10. S. Wold, H. Ruhe, H. Wold, and W. J. Dunn III, SIAM J. Sci. Sta. Comput. **5,** 735 (1984).
11. A. Höskuldsson, J. Chemometr. **2,** 211 (1988).
12. Q.-S. Sun, Z. Jin, P.-A. Heng, and D.-S. Xia, Lecture Notes in Computer Science **3686,** 268 (2005).
13. P. J. Besl and N. D. McKay, IEEE Trans. Pattern Anal. Mach. Intell. **14,** 239 (1992).
14. Y. Wang, F. Makedon, J. Ford, and H. Huang, in *Proceedings of the 26th Annual Int. Conf. IEEE EMBS* **4,** 2972 (2004).