Int. J. Electron. Commun. (AEÜ) 70 (2016) 427–435

Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

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

Image registration using a kernel partial least squares based mismatches removal method

Wei Zhao [a,*], Zheng Tian [a,b,c], Lijuan Yang [a], Weidong Yan [a], Jinhuan Wen [a]

[a] Departments of Applied Mathematics, Northwestern Polytechnical University, Xi'an 710129, China
[b] Department of Computer Science and Technology, Northwestern Polytechnical University, Xi'an 710129, China
[c] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, CAS, Beijing 100101, China

ARTICLE INFO

Article history:
Received 8 June 2015
Accepted 5 January 2016

Keywords:
Image registration
Feature matching
Kernel partial least squares (KPLS)
Mismatches removal

ABSTRACT

Feature point matching is a critical step in feature based image registration. For remote sensing images, scale invariant feature transform (SIFT) feature matching is often affected by similar descriptors and there are mismatches. To improve the quality of feature matching and image registration, we propose to use spatial relationship along with the SIFT descriptor for registration. Firstly, initial matches are obtained based on distances between SIFT feature descriptors. Secondly, the spatial relationship of matched points is encoded by kernel partial least squares (KPLS). By analyzing the collinearity of the KPLS features from coarse to fine, false matches are indicated. Finally, correct matches are used to realize accurate registration. Experimental results show an overall significant reduction of the mismatches while maintaining a high rate of correct matches. Compared with several other feature matching methods, the proposed method provides comparable or better results.

1. Introduction

For the past few decades, image registration is involved widely in many applications including image mosaic, change detection, image fusion, cartography, etc. [1]. The aim of image registration is to find the optimized transformation between the reference and sensed images, which differ in certain aspects, e.g., translation, scaling, rotation, and affine, but essentially contain overlapping scenes. The feature based registration methods consist of five steps: Preprocessing, feature detection, feature matching, transform model estimation, and resampling [2]. Feature detection and feature matching are two key steps among these operations. Some sophisticated feature detection methods have been applied successfully, such as SIFT [3], speeded up robust features (SURF) [4] and salient image disks (SID) [5,6]. However, because there are regions with similar intensity distribution in the reference and sensed images, the local descriptor based matching process often results in many mismatches which have negative effect on the estimation of transform model.

To address the issue, the robust feature matching methods with the ability of outliers rejection were researched in recent years. For SIFT based registration, Euclidean distance ratio filter (EDRF) [3] is often used to remove unreliable matches. EDRF excludes an unreliable match if the Euclidean distance ratio of closest to second-closest neighbors is less than a threshold. However, there are many mismatches if the threshold is high and many correct matches are excluded if the threshold is low. So Li et al. proposed scale-orientation joint restriction criteria to achieve robust feature matching [7]. They used the joint distance ratio filter to remove mismatches. The joint distance ratio filter rejects outliers according to the distance ratio between the first and second nearest-neighbors in the sense of joint distance that takes the scale and orientation information of SIFT key points into consideration. Similarly, a convenient and effective mode-seeking (MS) algorithm which also exploits the scale, orientation, and position information of SIFT features was presented in [8]. Hasan et al. proposed a two-step procedure to improve SIFT-based matching by taking advantage of neighborhood information [9]. They used EDRF and Random Sample Consensus (RANSAC) [10] to find primary matched feature points in the first step. Then EDRF is used in the neighborhoods of the primary matched feature points to find more secondary matched feature points in the second step.

Another way to overcome the problem is to combine the local descriptor based matching and the structure based matching. The structure based matching methods utilize the spatial relationship between feature points to establish the matching. For instance, the graph matching methods have been extensively used for locating correspondences between features [11–13] in computer

* Corresponding author. Tel.: +86 15291911446.
E-mail address: wu.kong1@163.com (W. Zhao).

http://dx.doi.org/10.1016/j.aeue.2016.01.001
1434-8411/

vision problems. Especially, Wang et al. discussed the relationship between graph matching and kernel principal component analysis (KPCA) [12]. From the perspective of KPCA, applying a similarity function to the original data set is equivalent to the process of using a kernel function to map the data into a higher, possibly infinite dimensional space. More recently, some robust point matching methods that considered the structure information have been proposed for computer vision and remote sensing image registration. Aguilar et al. [14] proposed a graph transformation matching (GTM) method which compares adjacent matrices of the nearest-neighbor graphs emerging from candidate matches to remove outliers. Liu et al. [15] used the spatial order of adjacent points to construct structure descriptors for feature points and determine candidate outliers. Similarly, Zhang et al. [16] presented a structure descriptor based on the triangle-area representation of the K nearest neighbors to determine the candidate outliers. They both used the change of the root mean square error (RMSE) before and after removing a candidate outlier to determine the true outliers. Shi et al. [17] make use of improved shape context in point structure description and GTM for exact matching. Yan et al. [18] proposed a false match rejection strategy based on the collinearity of the first pair of kernel canonical correlation analysis (KCCA) features. However, the above mentioned methods tend to obtain higher accuracy at the cost of reserving fewer matches. This leads to the removal of high quality matches, which suppresses the registration performance.

To reserve more high quality matches, RANSAC [10] is a classic and effective method to estimate the transformation parameters and establish correspondences, and it has been widely used in remote sensing image registration [19,20]. RANSAC is a nondeterministic method that it produces a reasonable result only with a certain probability. More iterations are needed to increase the probability. To reduce the randomness of RANSAC and keep more high quality matches, Moisan and Stival proposed Optimized Random Sampling Algorithm (ORSA) [21] and applied it to SIFT based image registration [22]. ORSA can find more correct matches than RANSAC with low randomness.

In this paper, a coarse to fine mismatches removal method based on KPLS is proposed for remote sensing image registration under the KPCA view of graph matching [12] and the collinearity property of KPLS features. The method is deterministic without randomness and can keep many high quality matches favorable to image registration. Its rationale is to reduce mismatches caused by similar descriptors with the help of the spatial relationship among matching features. By analyzing the collinearity of the KPLS features, one can decide whether there are outliers in the current matching with an acceptable tolerance. The rest of this paper is organized as follows: Section 2 reviews the related work of KPLS. The outliers removal strategy proposed is drawn in Section 3. Experiments and analysis aiming at evaluating the performance of the proposed method are detailed in Section 4, and conclusions are drawn in Section 5.

2. A review of kernel partial least squares technique

The partial least squares (PLS) method, which was initially developed by Wold et al. [23], has been a tremendously successful method for data analysis in the chemometrics and chemical industries [24]. Considering two centered multivariate random vectors in paired form $(\boldsymbol{x}, \boldsymbol{y})$ where $\mathbf{x} \in \mathbf{R}^{d_1}$ and $\mathbf{y} \in \mathbf{R}^{d_2}$, PLS focuses on studying the covariance between the two parts live in different spaces. Suppose that $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ and $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n)$ are n observations of $\mathbf{x}$ and $\mathbf{y}$ respectively, $\mathbf{M}_c = \mathbf{I}_n - (1/n)\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$ is the centering matrix where $\mathbf{I}_n \in \mathbf{R}^{n\times n}$ is an identity matrix and $\mathbf{1}_n = (1, 1, \cdots, 1)^{\mathrm{T}} \in \mathbf{R}^n$ is a column vector. The covariance matrix between $\mathbf{x}$ and $\mathbf{y}$ could be estimated by $\mathbf{C}_{xy} = (1/n)\mathbf{X}\mathbf{M}_c\mathbf{M}_c\mathbf{Y}^{\mathrm{T}}$. The aim of PLS is to find two directions $\mathbf{u} \in \mathbf{R}^{d_1}$ and $\mathbf{v} \in \mathbf{R}^{d_2}$, such that the covariance of the two projections $s_x = \mathbf{u}^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu}_x)$ and $t_y = \mathbf{v}^{\mathrm{T}}(\mathbf{y} - \boldsymbol{\mu}_y)$ are maximized, that is

$$\max_{u,v} cov(s_x, t_y) = \max_{u,v} \mathbf{u}^{\mathrm{T}}\mathbf{C}_{\mathbf{xy}}\mathbf{v}, \quad s.t. \quad \mathbf{u}^{\mathrm{T}}\mathbf{u} = \mathbf{v}^{\mathrm{T}}\mathbf{v} = 1. \tag{1}$$

Then $\mathbf{u}$ and $\mathbf{v}$ are the left and right singular vectors of $\mathbf{C}_{\mathbf{xy}}$, respectively, according to the Lagrangian multiplier method. The directions of maximal covariance are given by all the singular vectors of $\mathbf{C}_{\mathbf{xy}}$ according to [25].

KPLS is the nonlinear extension of PLS using the "kernel trick" and PLS is in fact a special case of KPLS with linear kernel. Let $\boldsymbol{\Phi}_1 = (\phi_1(\mathbf{x}_1), \phi_1(\mathbf{x}_2), \cdots, \phi_1(\mathbf{x}_n))$ and $\boldsymbol{\Phi}_2 = (\phi_2(\mathbf{y}_1), \phi_2(\mathbf{y}_2), \cdots, \phi_2(\mathbf{y}_n))$ denote the corresponding images of the data matrices $\mathbf{X}$ and $\mathbf{Y}$ in Hilbert space. Their kernel matrices are $\mathbf{K}_{\mathbf{X}}$ and $\mathbf{K}_{\mathbf{Y}}$, where the elements are computed by the kernel function

$$\mathbf{K}_{\mathbf{X}}(i,j) = k_x(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi_1(\mathbf{x}_i), \phi_1(\mathbf{x}_j) \rangle, \tag{2}$$

$$\mathbf{K}_{\mathbf{Y}}(i,j) = k_y(\mathbf{y}_i, \mathbf{y}_j) = \langle \phi_2(\mathbf{y}_i), \phi_2(\mathbf{y}_j) \rangle, \tag{3}$$

with $1 \le i, j \le n$. Denote $\bar{\boldsymbol{\Phi}}_1 = \boldsymbol{\Phi}_1\mathbf{M}_c$ and $\bar{\boldsymbol{\Phi}}_2 = \boldsymbol{\Phi}_2\mathbf{M}_c$ are the centered version of $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$, respectively. Then the corresponding kernel matrices become $\bar{\mathbf{K}}_{\mathbf{X}} = \mathbf{M}_c\mathbf{K}_{\mathbf{X}}\mathbf{M}_c$ and $\bar{\mathbf{K}}_{\mathbf{Y}} = \mathbf{M}_c\mathbf{K}_{\mathbf{Y}}\mathbf{M}_c$. Suppose $\mathbf{u}$ and $\mathbf{v}$ can be linearly expressed by $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$, that is $\mathbf{u} = \bar{\boldsymbol{\Phi}}_1\boldsymbol{\alpha}$ and $\mathbf{v} = \bar{\boldsymbol{\Phi}}_2\boldsymbol{\beta}$, the optimization problem of Eq. (1) could be reformulated as

$$\max_{\alpha,\beta} \boldsymbol{\alpha}^{\mathrm{T}}\bar{\mathbf{K}}_{\mathbf{X}}\bar{\mathbf{K}}_{\mathbf{Y}}\boldsymbol{\beta}, \quad s.t. \boldsymbol{\alpha}^{\mathrm{T}}\bar{\mathbf{K}}_{\mathbf{X}}\boldsymbol{\alpha} = 1, \boldsymbol{\beta}^{\mathrm{T}}\bar{\mathbf{K}}_{\mathbf{Y}}\boldsymbol{\beta} = 1. \tag{4}$$

According to the Lagrangian multiplier method, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the eigenvectors of the generalized eigenvalue problem

$$\begin{pmatrix} 0 & \bar{\mathbf{K}}_{\mathbf{X}}\bar{\mathbf{K}}_{\mathbf{Y}} \\ \bar{\mathbf{K}}_{\mathbf{Y}}\bar{\mathbf{K}}_{\mathbf{X}} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \lambda \begin{pmatrix} \bar{\mathbf{K}}_{\mathbf{X}} & 0 \\ 0 & \bar{\mathbf{K}}_{\mathbf{Y}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix}. \tag{5}$$

Suppose μ_x and μ_y are the mean vectors of $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$, respectively. Once the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are determined by (5), given any $\mathbf{x}$ and $\mathbf{y}$, the projections are computed as

$$s_x = \langle \phi_1(\mathbf{x}) - \boldsymbol{\mu}_x, \mathbf{u} \rangle = \langle \phi_1(\mathbf{x}) - \boldsymbol{\mu}_x, \boldsymbol{\Phi}_1\boldsymbol{\alpha} \rangle = \bar{\mathbf{k}}_{\mathbf{x}}\boldsymbol{\alpha}, \tag{6}$$

$$t_y = \langle \phi_2(\mathbf{y}) - \boldsymbol{\mu}_{\mathbf{y}}, \mathbf{v} \rangle = \langle \phi_2(\mathbf{y}) - \boldsymbol{\mu}_{\mathbf{y}}, \boldsymbol{\Phi}_2\boldsymbol{\beta} \rangle = \bar{\mathbf{k}}_{\mathbf{y}}\boldsymbol{\beta}, \tag{7}$$

where $\bar{\mathbf{k}}_{\mathbf{x}}$ and $\bar{\mathbf{k}}_{\mathbf{y}}$ are the rows of $\bar{\mathbf{K}}_{\mathbf{X}}$ and $\bar{\mathbf{K}}_{\mathbf{Y}}$ corresponding to $\phi_1(\mathbf{x}) - \boldsymbol{\mu}_{\mathbf{x}}$ and $\phi_2(\mathbf{y}) - \boldsymbol{\mu}_{\mathbf{y}}$, respectively. A linear regression model could be estimated based on the paired point sets $\{(s_i, t_i), i = 1, 2, \cdots, n\}$. The projections $\{(s_i, t_i), i = 1, 2, \cdots, n\}$ are called the KPLS features.

3. Mismatches removal method based on KPLS

Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ and $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n)$ be the position vectors of one-to-one correspondence pairs of matching. $\boldsymbol{\Phi}_1 = (\phi_1(\mathbf{x}_1), \phi_1(\mathbf{x}_2), \cdots, \phi_1(\mathbf{x}_n))$ and $\boldsymbol{\Phi}_2 = (\phi_2(\mathbf{y}_1), \phi_2(\mathbf{y}_2), \cdots, \phi_2(\mathbf{y}_n))$ denote their corresponding images in Hilbert space. At first $(s_i, t_i), i = 1, 2, \cdots, n$ should be computed. However, because $\bar{\mathbf{K}}_{\mathbf{X}}$ and $\bar{\mathbf{K}}_{\mathbf{Y}}$ are often singular, Eq. (5) is not used but the KPCA dimensionality reduction is carried out on $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ to reduce the singularity alternatively. Suppose the dimensions of $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ are reduced to r and their KPCA representations are $\boldsymbol{\Psi}_1 = (\psi_1(\mathbf{x}_1), \psi_1(\mathbf{x}_2), \cdots, \psi_1(\mathbf{x}_n))$ and $\boldsymbol{\Psi}_2 = (\psi_2(\mathbf{y}_1), \psi_2(\mathbf{y}_2), \cdots, \psi_2(\mathbf{y}_n))$, respectively. In the ideal case that $r = n$, using the KPCA representations and Eq. (1) to calculate $(s_i, t_i), i = 1, 2, \cdots, n$ is the same as KPLS because KPCA is an orthogonal transformation in Hilbert space. When $r < n$, the method gives the approximate solution. KPCA preprocessing is in fact structure encoding according to [12].