Canonical Correlation Analysis Applied to Remove SIFT mismatching

Weidong Yan, Zheng Tian and Xifa Duan
School of Science,
Northwestern Polytechnical University,
Xi'an, China
weidongyan@qq.com

Abstract—**In this paper, we propose a novel remote sensing image registration method based on the combination of local Scale Invariant Feature Transform (SIFT) description and the spatial relationship of matching features, which is accomplished based on Canonical Correlation Analysis (CCA). Compared with SIFT which is often impacted by similar structures, the retrofitted SIFT algorithm is more robust and accurate. The method proceeds in two stages. In the first stage a putative set of correspondences is obtained based on distances between SIFT feature descriptors. In the second stage the matches are refined by imposing spatial relationship of matching features by means of CCA and the incorrect matches are rejected as outliers. By employing the CCA algorithm, a more robust and accurate registration result is achieved at the expense of moderate computational complexity. Experimental results show an overall significant reduction of the mismatches while maintaining a high rate of correct matches.**

Keywords-Canonical Correlation Analysis (CCA); Mismatching; Image registration; Scale Invariant Feature Transform (SIFT)

I. INTRODUCTION

For decades, image registration has drawn many researchers' attention and has become a hot research topic. It has been widely used in computer vision applications, such as image mosaic, change detection, cartography, and so on. The primary objective of image registration is to match two or more images that differ in certain aspects, e.g., translation, scaling, and rotation, but essentially represent the same scene [1]. Existed registration approaches can be generally categorized into two major categories: area-based and feature-based methods [2]. Although many methods have been already proposed for different applications, it is still a challenging task to develop a robust image registration algorithm for the following two main reasons [3]. First, there always exist geometric distortions stemmed from rotation, scale, affine transformations, etc. Second, duo to features in an image may partially appear or even disappear at all in another one and the existing of unavoidable noise will result in outliers in the estimation set.

One feasible way to solve the first issue is to extract features from reference and sensed images separately, followed by picking a descriptor to find out an appropriate relation between the two sets of features [4]. If the descriptor is independent, stable, and invariant to image transformation, then the image registration will attain a reliable performance [5]. During the last decades, many feature-based methods devote to extracting features such as point, line, region, and descriptor [6]. Among them, the SIFT descriptor [7] has been demonstrated to be a potential robust alternative for point feature-based registration [8]. Some improvements have been suggested for the standard SIFT, such as PCA-SIFT [9], robust SIFT [10, 11], and also those in [12]-[14]. Despite the fact that great progress has been made in the past, sometimes matches obtained by comparing feature descriptors are often not reliable. Global geometric constraints are usually used to identify true correspondences.

To address the second issue, it is necessary to develop outlier removal technique to enhance the robustness and improve the accuracy of image registration. Wong and Clausi [4] presented a method, which uses an improved random sample consensus (RANSAC) algorithm to match the feature points. It works by selecting random subsets of feature points from which a tentative geometric transformation is computed. If the transformation consistently extends to a significant portion of the full set, then it is accepted as correct. Wang and Zhang [15] employ the geometrical and topological relationship among putative matches to discard mismatches, but ignore the scale and orientation information of the potential feature pairs, which can express a similarity transformation.

Considering the aforementioned factors involved in image registration, we thus propose a novel image registration approach based on SIFT and CCA. The rationale of the proposed approach is to reduce false matches caused by similar structures with the help of the spatial relationship of matching features. The main contributions of the proposed algorithm include the following: 1) we obtain a higher correct matching rate by retrofitting the SIFT algorithm; 2) we propose an effective criterion to identify the mismatches; 3) with the help of the algorithm, the accuracy of the image registration is greatly improved. Finally, we show through experiments that CCA-based algorithms consistently provide better performance than standard RANSAC method.

This work was supported by the National Natural Science Foundation of China (No. 61201323, 60972150, 11172235), Northwestern Polytechnic University Foundation of Fundamental Research (No.JC20110277).

II. PRINCIPLE OF SIFT AND CCA

A. Scale Invariant Feature Transform

In computer vision systems and pattern recognition, feature descriptors extracted from an image's gray values are usually used. Scale invariant feature transform (SIFT) is one of the best descriptors for feature matching [7]. The SIFT algorithm transforms image data into scale-invariant coordinates relative to local features and is based on four major stages:

- Scale-space extrema detection: The image is first convolved with a series of Gaussian filters at different scales. Then, adjacent Gaussian images are subtracted to produce the difference-of-Gaussian images. Scale-space extrema in the difference-of-Gaussians are regarded as the most stable scale-invariant features.
- Determination of keypoint location: Scale-space extrema are interpolated to obtain subpixel accuracy. Candidate keypoints with low contrast and those that are located along an edge but unstable to small amounts of noise are eliminated.
- Orientation assignment: This stage is the orientation assignment to each keypoint, based on local image gradient directions. This allows for the representation of each keypoint relative to this orientation, achieving invariance to image rotation. Peaks in orientation histogram are supposed to be dominant directions.
- Keypoint descriptor assignment: The previously described steps assigned the location, scale, and orientation of each keypoint. The motivation for the computation of a more complex descriptor is to obtain a highly distinctive keypoint and invariant as possible to variations. Each resultant SIFT descriptor is a 128-element feature vector.

After the keypoint descriptor has been calculated, keypoints are matched by using the minimum distance method, where an exhaustive search between all keypoints in both images is performed. In order to increase the stability of matching results, the ratio between the distance of the closest neighbor and the distance to the second closest neighbor is calculated to reject the matches. More details about SIFT can be referred to [7]. Despite the outstanding characteristics of the SIFT, it has some problems with image registration. Even after the identification of matching candidates after removal of incorrect initial matches as described above, there are still many false matches due to feature points located in some similar structures, which lead to a further outlier removal.

B. A review of Canonical Correlation Analysis Technique

Proposed by H.Hotelling in 1936 [16], Canonical Correlation Analysis learns two maximally correlated subspace from two datasets without explicitly controlling the shared basis vectors and require two datasets such that each example in the first dataset correspond to one example in the second dataset. Recently, it has been applied to a variety of computer vision and pattern recognition problems, including dimension reduction and image retrieval. In CCA two different representations of the same set of objects are given, and a projection is computed for each representation such that they are maximally correlated in the dimensionality-reduced space.

Given a sample from a pair dataset $\{(x_1, y_1), \cdots, (x_n, y_n)\}$, $x_i \in \mathbb{R}^2$ and $y_i \in \mathbb{R}^2$ denote the coordinates of the i th matching pairs. Let $X = [x_1, x_2, \cdots, x_n] \in \Re^{2\times n}$ and $Y = [y_1, y_2, \cdots, y_n] \in \Re^{2\times n}$ be data matrices. We assume that both $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ are centered, i.e. $\sum_{i=1}^n x_i = 0, \sum_{i=1}^n y_i = 0$. CCA simultaneously finds basis vectors w_1 and c_1, one for X and the other for Y, such that the correlations between the projections of the variables onto these basis vectors are mutually maximized. This is expressed as

$$\rho = \max_{w_1, c_1} \frac{\mathrm{cov}(w_1^T x, c_1^T y)}{\sqrt{D(w_1^T x) D(c_1^T y)}} = \max_{w_1, c_1} \frac{w_1^T XY^T c_1}{\sqrt{w_1^T XX^T w_1 c_1^T YY^T c_1}} \tag{1}$$

The maximum canonical correlation is the maximum of ρ with respect to w_1 and c_1. The CCA optimization problem formulated in equation (1) is equivalent to the following optimization problem with the constraints:

$$\begin{aligned} &\max_{w_1, c_1} w_1^T XY^T c_1 \\ &s.t.\ w_1^T XX^T w_1 = c_1^T YY^T c_1 = 1 \end{aligned} \tag{2}$$

This optimization problem can be solved by the Lagrange multiplier method [17], resulting in the following generalized eigenproblem

$$\begin{pmatrix} 0 & XY^T \\ YX^T & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ c_1 \end{pmatrix} = \lambda \begin{pmatrix} XX^T & 0 \\ 0 & YY^T \end{pmatrix} \begin{pmatrix} w_1 \\ c_1 \end{pmatrix} \tag{3}$$

with w_1 and c_1 as eigenvectors. Once the basis vector pair (w_1, c_1) is found, a second pair of basis vector (w_2, c_2) can be obtained that generates components which are maximally correlated with each other and uncorrelated with the components of the first pair. The next pair(s) of basis vectors can again be found by maximizing (1), restricted to the fact that the components are uncorrelated to all previously found components.

As mentioned above, CCA is appropriate to model a pair of related data sources that provide two different views of the same object. So, we present an outlier removal strategy with the help of CCA.

III. A NEW OUTLIER REMOVAL STRATEGY

A. *Criterion of Outlier Rejection*

In section 2.2, we have acquired basis vector pair (w_1, c_1). Therefore, for all $x_i \in X, y_i \in Y$, then the canonical correlation features can be extracted as follows:

$$F_1 = w_1^T X \ , \ G_1 = c_1^T Y \tag{4}$$

where $F_1(i) = w_1^T x_i$ and $G_1(i) = c_1^T y_i$.

Let's assume that the relationship between X and Y is a linear one. According to CCA, F_1 and G_1 is a linear relationship. Given points $(F_1(i), G_1(i))$, we can fit a line $ax + by + c = 0$. The points $(F_1(i), G_1(i))$ deviate from the line by different amounts. "Good" points have a small deviation; "Bad" points have a large deviation.

Definition: (Influence Function) Given a point (x_0, y_0) and the equation of a line $ax + by + c = 0$, the distance from the point to the line is $\mathrm{dist}(i) = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}$. We can define normalized distance as influence function:

$$\mathrm{IF}(i) = \frac{\mathrm{dist}(i)}{\sum_{i=1}^{n} \mathrm{dist}(i)} = \frac{|aF(i) + bG(i) + c|}{\sum_{i=1}^{n} |aF(i) + bG(i) + c|} \tag{5}$$

Criterion of Outlier Rejection: The obtained influence function and a preset threshold T are compared to exclude false matches. If $\mathrm{IF}(i) > T$, then x_i and y_i are mismatch. One the other hand, if $\mathrm{IF}(i) \le T$, then this indicate that x_i and y_i are matching.

B. *Outlier Removal Strategy*

The following summary of the outlier rejection algorithm is provided to list the key steps of the process, and the order in which they were carried out.

Algorithm : The outlier rejection procedure by CCA

Stage Ⅰ: Performed initial matching using SIFT

(1) Computation of the SIFT descriptors of the reference and sensed images;

(2) Establish initial matching relation by computing Euclidean distance between all potential matching pairs;

Stage Ⅱ: Fit a line

(1) Construct CCA model using position information of initial matching pairs, and calculate canonical correlation features (F_1, G_1);

(2) Fit a line $ax + by + c = 0$ using $(F_1(i), G_1(i))$;

Stage Ⅲ: Performed outlier rejection using CCA

(1) Compute influence function

$$\mathrm{IF}(i) = \frac{|aF(i) + bG(i) + c|}{\sum_{i=1}^{n} |aF(i) + bG(i) + c|},$$

(2) If the distance between the candidate match i and line L is larger than a threshold, i.e., $\mathrm{IF}(i) \ge T$, then match i is a mismatch and is rejected. If the $\mathrm{IF}(i)$ is below T, the match i is validated, where T is a preset threshold.

(3) Estimate the geometric model using refined matching points, and implement image registration.

In our experiments, the threshold T can be computed as the mean influence:

$$T = \frac{1}{n} \sum_{i=1}^{n} \mathrm{IF}(i) \tag{6}$$

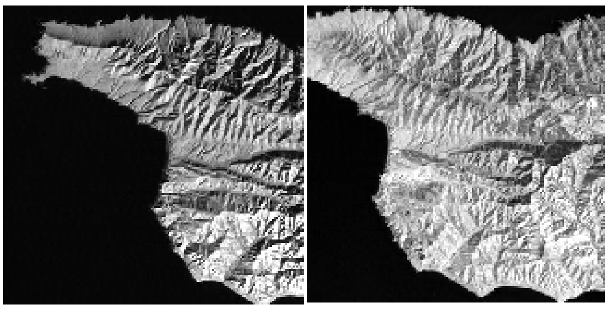
IV. EXPERIMENTS AND DISCUSSION

For experimental verification, two multispectral (Landsat TM) images with different look angles are used in this experiment (Fig.1 (a) and (b)). The images show the scene on different dates. They are provided by University of California at Santa Barbara Vision Laboratory. The traditional matching technique based purely on the SIFT descriptor similarity is too fragile. As shown in Fig. 1(c), some initial matches are false, such as the matches of 2, 45 and 73. While with the help of the influence function provided by the CCA, the proposed approach can reject such false matches easily. By calculation, the best fit straight line is

$$L: 0.9641x - y - 0.6517 = 0 \tag{7}$$

which is shown in Fig.1(d) in red, and the mean influence is $T = 0.04$.

As illustrated in Fig. 1(e), we observe that the matches of 2, 45 and 73 are false matches, and the values of influence function are IF(2)=1.1798, IF(45)=1.202 and IF(73)=1.0573, respectively. Those values are greater than T, in other words, the candidate matches (2, 45 and 73) is far from the line L (see Fig. 1(d)). From the results, we observe that the use of spatial relationship of matching features with our approach can remove correspondences that do not satisfy the defined constraints.



(a) (b)

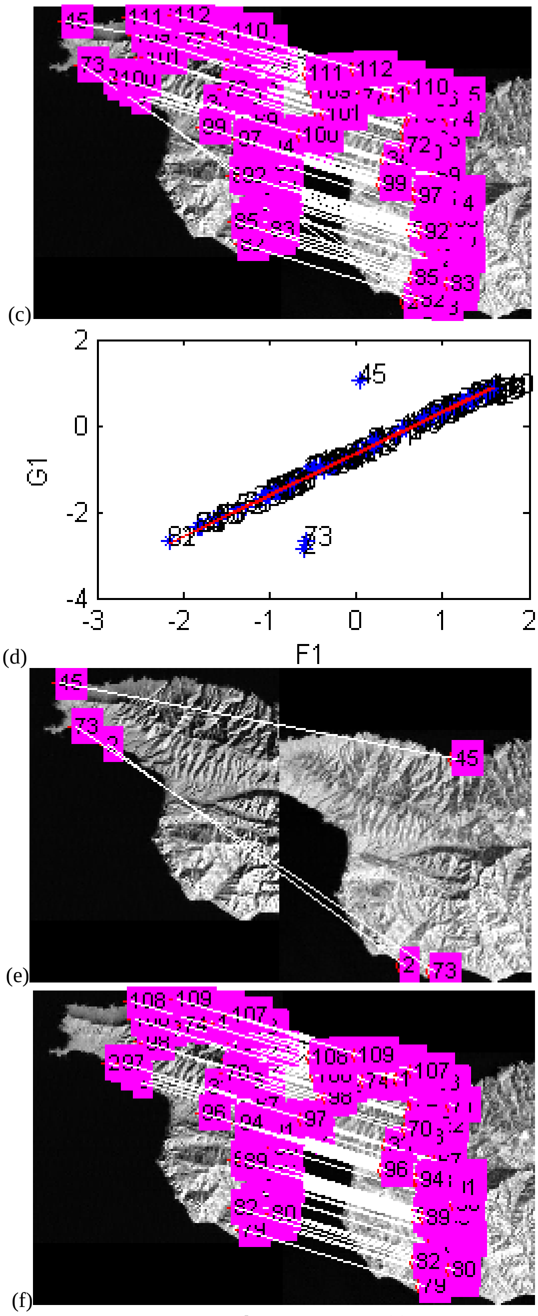


Figure 1. Image registration on Landsat TM images. (a) TM image in 1984; (b) TM image in 1986; (c) matches found using SIFT; (d) scatter plot plus fitted straight line; (e) mismatches found using CCA; (f) correct matches.

To compare the registration performance, we commence our real-world evaluation of the feature matching method on remote sensed images. In this experiment, we use the following measures to evaluate the effectiveness of the proposed approach: the number of correct matches, the number of false matches, and the registration accuracy. Experiments have been carried out on two remote sensing image pairs comparison with the following two state-of-the-art feature matching methods: Scale Invariant Feature Transform (SIFT) [7] and RANSAC [4]. Fig. 2 shows registration results of two remote sensing image pairs with affine transformation. The two images in the first pairs are optical remote sensing images (Tokyo Bay, Japan) but acquired by different sensors. Fig. 2(a) was obtained by Landsat TM, and Fig. 2(b) was obtained by ASTER sensor, which are used as the reference image and the sensed image, respectively. The matching results based on SIFT descriptor similarity for the two images are shown in Fig. 2(c), and the mismatches found by CCA is shown in Fig. 2(d). The registration result is shown in Fig. 2(e).

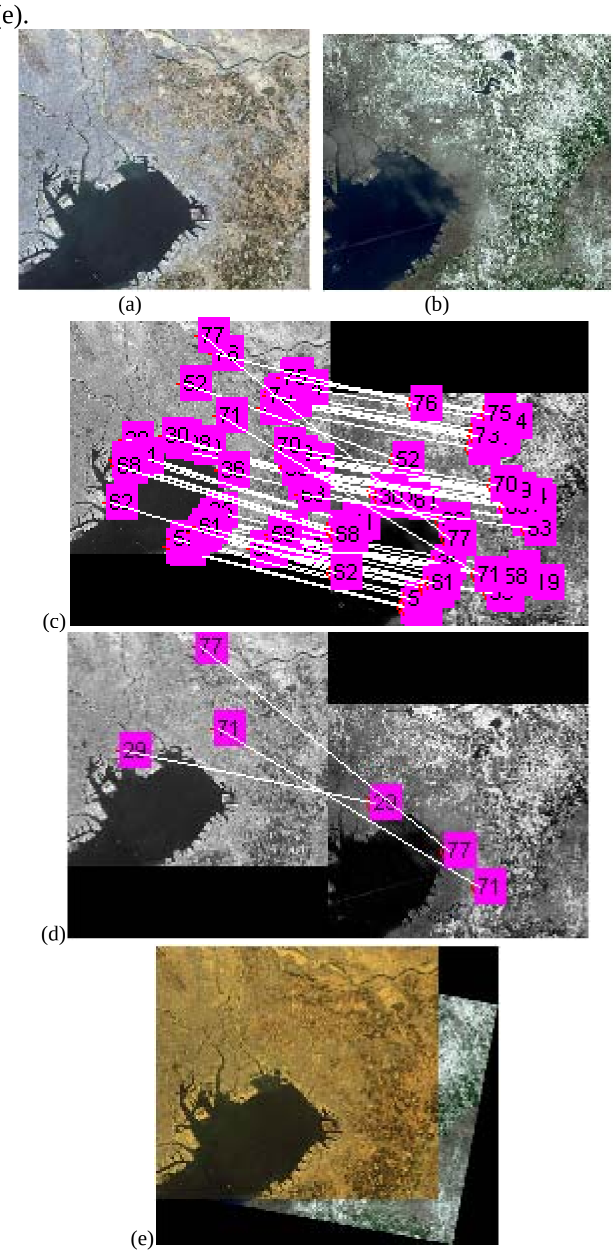


Figure 2. Image registration on Landsat TM images. (a) Reference image; (b) Sensed image; (c) matches found using SIFT; (d) mismatches found using CCA; (e) The registration result.

TABLE Ⅰ Matching results for Figs. 1, 2 and 3 obtained by the three methods

Image	CCA-Based		RANSAC-Based		SIFT	
	Correct matches	False matches	Correct matches	False matches	Correct matches	False matches
Fig.1	109	0	91	0	109	3
Fig.2	74	0	70	0	74	3

After checking by hand, the number of correct and false matches is summarized in Table 1. From these experiments, it is clear that our CCA-based approach gives encouraging results when compared with the approaches of RANSAC.

To evaluate the performance of all three procedures, we consider the root mean square error (RMSE) between the corresponding point pairs. RMSE is the conventional and most widely used measure. When evaluating an estimator $\hat{\mathrm{T}}$ of the geometrical transformation T, it is defined to be

$$\mathrm{RMSE} = \sqrt{\frac{1}{n^2}\sum_{i=1}^{N}(\hat{\mathrm{T}}(x_i) - y_i)^2} \tag{8}$$

where $\hat{x}_i = \hat{\mathrm{T}}(x_i)$ are the estimated coordinates and y_i are the coordinates of the ith corresponding point pair, N is the number of corresponding point pairs. If the value of RMSE is smaller, then the registration is regarded better. The registration accuracies in terms of the RMSE for Fig. 1, 2 and 3 by using the three different methods are summarized in Table 2. From the Table 2, it can be seen that the proposed algorithm is better than all two competing algorithms in terms of RMSE, which show our method performs best in the three different methods.

TABLE Ⅱ The accuracy (RMSE) of registration for Figs. 1, 2 and 3 obtained by the three methods

Image	CCA-Based	RANSAC-Based	SIFT
Fig.1	0.7624	2.0033	58.8365

V. CONCLUSIONS

In this paper, a mismatching elimination algorithm, inspired by CCA, has been proposed to improve the accuracy of the SIFT-based method for the image registration problem. This algorithm is applicable to various sensing images with different sensors, acquisition times, and scene changes. The core of the proposed method was to impose spatial relationship of matching features by means of CCA to reject outliers. The experimental results on a variety of multisource remote sensing image pairs proved the method’s advantages in terms of robustness and accuracy. Despite the effectiveness of the proposed approach, many developments need to be considered in the future work, including the choice of threshold and the algorithm efficiency.

REFERENCES

[1] M. Xu, P.K. Varshney, “A Subspace Method for Fourier-Based Image Registration,” *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 491- 494, 2009.

[2] Barbara Zitova, Jan Flusser, “Image registration methods: A survey,” *Image Vis. Comput.*, vol. 21, no. 11, pp. 977-1000, 2003.

[3] Zhili Song, Junping Zhang, “Remote Sensing Image Registration Based on Retrofitted SURF Algorithm and Trajectories Generated From Lissajous Figures,” *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 491-495, 2010.

[4] A. Wong, D. A. Clausi, “ARRSI: Automatic registration of remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1483-1493, 2007.

[5] K. Mikolajczyk, C. Schmid, “A performance evaluation of local descriptors,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615-1630, 2005.

[6] K. Mikolajczyk, C. Schmid, “Scale and affine invariant interest point detectors,” *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63-86, 2004.

[7] D. Lowe, “Distinctive image features from scale invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91-110, 2004.

[8] P. Schwind, S. Suri, P. Reinartz, and A. Siebert, “Applicability of the SIFT operator to geometric SAR image registration,” *Int. J. Remote Sens.*, vol. 31, no. 8, pp. 1959-1980, 2010.

[9] Y. Ke, R. Sukthankar, “PCA-SIFT: A more distinctive representation for local image descriptors,” *in Proc. Conf. Comput. Vis. Pattern Recog.*, Washington, DC, 2004, pp. 506-513.

[10] Q. Li, G. Wang, J. Liu, and S. Chen, “Robust scale-invariant feature matching for remote sensing image registration,” *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 287-291, 2009.

[11] Amin Sedaghat, Mehdi Mokhtarzade, and Hamid Ebadi, “Uniform Robust Scale-Invariant Feature Matching for Optical Remote Sensing Images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4516-4527, 2011.

[12] P. Moreno, A. Bernardino, and J. S. Victor, “Improving the SIFT descriptor with smooth derivative filters,” *Pattern Recognit. Lett.*, vol. 30, no. 1, pp. 18-26, 2009.

[13] A. Lingua, D.Marenchino, and F. Nex, “Performance analysis of the SIFT operator for automatic feature extraction and matching in photogrammetric applications,” *Sensors*, vol. 9, pp. 3745-3766, 2009.

[14] Shanhu Wang, Hongjian You, Kun Fu, “BFSIFT: A Novel Method to Find Feature Matches for SAR Image Registration,” *IEEE Geosci. Remote Sensing Lett.*, vol. 9, no. 4, pp. 649-653, 2012.

[15] Y.T. Wang, D.Z. Zhang, and J.W. Tian, “Discarding wide baseline mismatches via topological clustering,” *Electron. Lett.*, vol. 44, no. 11, pp. 670-671, 2008.

[16] Harold Hotelling, “Relations between two sets of variates,” *Biometrika*, vol. 28, no. 3-4, pp. 321-377, 1936.

[17] Hardoon, D. R., S. Szedmak, J. Shawe-Taylor, “Canonical Correlation Analysis: An Overview with Application to Learning Methods,” *Neural Computation*, vol. 16, no. 12, pp. 2639-2664, 2004.