A Novel Semi-supervised Classification Method Based on Class Certainty of Samples

Fei Gao[1], Zhenyu Yue[1](✉), Qingxu Xiong[1], Jun Wang[1], Erfu Yang[2], and Amir Hussain[3]

[1] Beihang University, Beijing 100191, China
feigao2000@163.com, zhenyu_yue@163.com
[2] University of Strathclyde, Glasgow G1 1XJ, UK
[3] University of Stirling, Stirling FK9 4LA, UK

Abstract. The traditional classification method based on supervised learning classifies remote sensing (RS) images by using sufficient labelled samples. However, the number of labelled samples is limited due to the expensive and time-consuming collection. To effectively utilize the information of unlabelled samples in the learning process, this paper proposes a novel semi-supervised classification method based on class certainty of samples (CCS). First, the class certainty of unlabelled samples obtained based on multi-class SVM is smoothed for robustness. Then, a new semi-supervised linear discriminant analysis (LDA) is presented based on class certainty, which improves the separability of samples in the projection subspace. Finally, the nearest neighbor classifier is adopted to classify the images. The experimental results demonstrate that the proposed method can effectively exploit the information of unlabelled samples and greatly improve the classification effect compared with other state-of-the-art approaches.

Keywords: Remote sensing images · Semi-supervised classification
Class certainty · Semi-supervised LDA

1 Introduction

With the rapid development of the remote sensing (RS) technology, the higher-resolution and more informative RS images can be acquired, and are already used in target surveillance, disaster relief, environmental protection and *etc.* [1–3]. The process of RS images interpretation consists of three parts: target detection, image segmentation and image classification [4, 5]. Besides, image classification is the most critical step. However, since the sample labeling for RS image is time-consuming, it's difficult to achieve accurate classification of RS images when the labelled samples are insufficient, which has become one of research hotspots [6, 7].

Generally, the working mechanism of human cognitive system have inspired researchers to improve the classification accuracy of images with insufficient labelled samples. Since most of the information received by the brain is unlabelled, the human cognition is a semi-supervised learning process, where the unlabelled information is utilized based on the priori knowledge. Inspired by this, many semi-supervised learning

J. Ren et al. (Eds.): BICS 2018, LNAI 10989, pp. 315–324, 2018.
https://doi.org/10.1007/978-3-030-00563-4_30

methods have been presented, such as generative mode, semi-supervised SVM [8], graph-based model [9], self-training model and co-learning model [10]. For the semi-supervised algorithms, unlabelled samples are used to enlarge initial labelled samples set and make the classification surface pass through the space with sparse samples. In [11], the transductive support vector machine (TSVM) is developed to search the optimal classification surface based on margin maximization by iteratively assigning the sample positive label or negative one. Persello and Bruzzone present a progressive semi-supervised SVM with diversity (PS3VM-D) to make candidate samples within and closer to the margin band [12]. Then, samples are incrementally selected among the candidates considering the kernel cosine-angular similarity. Based on co-training model, Zhou designed a tri-training algorithm by training three classifiers. Then, reliable unlabelled samples are selected by one classifier and added to the labelled samples set of the other two classifiers in an iterative way [13]. Although the aforementioned algorithms are proved to be effective experimentally, semi-supervised learning methods are not always helpful because of the strict requirements of data distribution, selection method and labeling method for unlabelled samples.

To effectively improve RS images classification performance, this paper proposes a novel semi-supervised classification method by utilizing unlabelled samples based on class certainty of samples (CCS). Different from other semi-supervised algorithms, CCS initially assigns the class certainty to unlabelled samples and integrates it to the scatter matrixes of linear discriminant analysis (LDA). The new scatter matrixes can effectively describe the true characteristic distribution, which makes samples more separable in the projection subspace. Since the class certainty is used to measure the class reliability of samples, the unlabelled samples with high reliability play a more important role than those with low reliability in CCS. To ensure the sufficient class reliability of unlabelled samples in the subsequent semi-supervised process, the class certainty is smoothed through normalization and threshold considering the complicated distribution of samples. As a result, the performance of CCS is greatly improved.

The rest of this paper is organized as follows. Section 2 describes the proposed method in detail. The experiments for the SAR targets classification are provided in Sect. 3 and the conclusions are drawn in Sect. 4.

2 Proposed Method

In this part, we first present the related definition. The training samples $X = [L, U] \in R^{d\times N}$ are divided into two parts according to the label of samples. Let $L = [x_1, x_2, \cdots, x_l] \in R^{d\times l}$ be the feature matrix of labelled samples with label vector $[y_1, y_2, \cdots, y_l]$, $y_i \in \{1, 2 \cdots, k\}$ and $U = [x_{l+1}, x_{l+2}, \cdots, x_{l+u}] \in R^{d\times u}$ be the feature matrix of unlabelled samples. $N = l + u$ denotes the number of training samples and the test set is T. Then, as shown in Fig. 1, the proposed novel semi-supervised method (CCS) consists of four main ingredients.

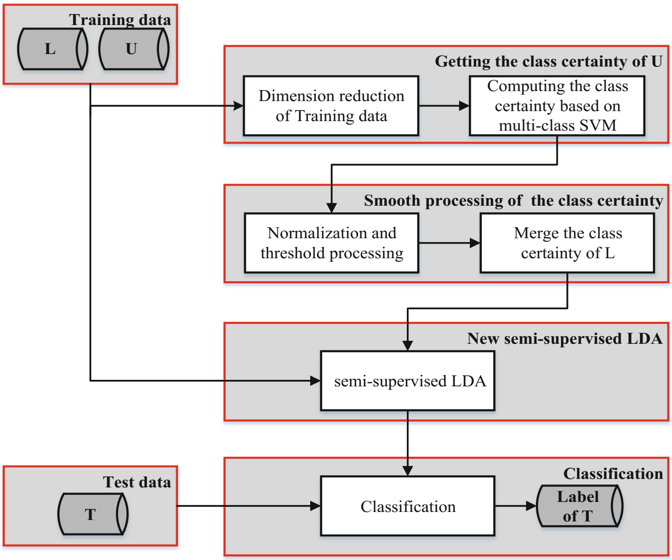


Fig. 1. Flowchart of the proposed method CCS.

2.1 Getting the Class Certainty of Unlabelled Samples

Dimension Reduction of Training Data. In Fig. 1, the inputs of CCS are the original RS training data L and U in high-dimension. To get the class certainty information, the computational complexity and the dimension of training data should be reduced. Thus, based on KLDA algorithm, the projection characteristics L_1 and U_1 are obtained.

Computing the Class Certainty Based on Multi-class SVM. The output of SVM can effectively measure the class certainty of samples. After the dimension reduction, SVM can be trained based on the labelled samples L_1. Because the samples are generally multi-class, we construct multi-class SVM based on the "one-against-one" approach. To express more clearly, an example of obtaining the class m certainty of unlabelled samples is shown in Fig. 2.

In Fig. 2, L_1^m denotes labelled samples of class m with reduced dimension. Let L_1^m and samples of the other classes be positive labels and negative labels, respectively. Then, after training k − 1 binary SVM between L_1^m and other classes of samples, the corresponding output vector is derived by passing U_1 to every binary SVM. For example, $f^{m,m-1} = (w^T\phi(U_1) + b)^T$ denotes the output vector of the SVM trained by L_1^m and L_1^{m-1}. To get the class m certainty f_U^m, the output vectors are added based on the voting method. Similarly, other class certainty f^i, $i = 1, 2, \cdots, k$, $i \neq m$ can be obtained according to the corresponding implementation.

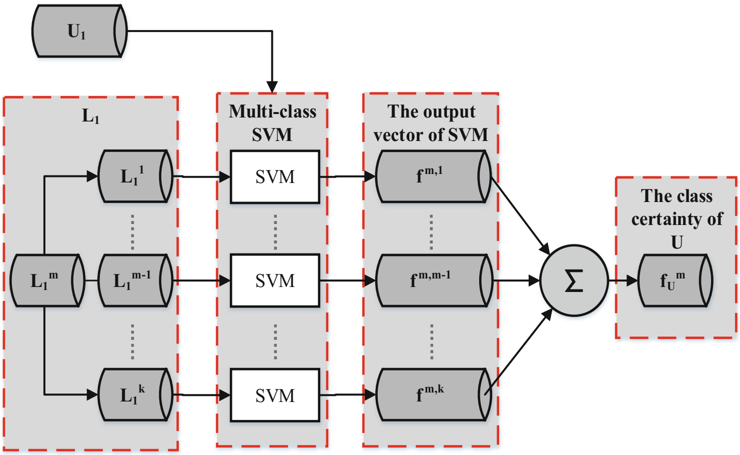


Fig. 2. Flowchart of obtaining the class m certainty of unlabelled samples.

2.2 Smooth Processing of the Class Certainty

Normalization and Threshold Processing. Since the class i certainty $f_U^{\mathrm{i}}(i \in \{1, 2, \cdots, k\})$ contains elements ranged from less than 0 to larger than 1, they should be normalized and threshold processed before utilizing. Accordingly, we utilize the min-max standard method, which can be written as

$$p_U^i = \frac{f_U^i - \min}{\max - \min}, \; i \in \{1, 2, \cdots, k\} \tag{1}$$

where p_U^i represents the class i certainty of U after normalization processing. Then, we choose threshold $t \in [0, 1]$. If the element of p_U^i is less than t, we set it to 0.

$$p_{U,j}^i = \begin{cases} 0, & p_{U,j}^i < t \\ p_{U,j}^i, & others \end{cases}, \; i \in \{1, 2 \cdots, k\}, \; others\, j \in \{1, 2, \cdots, u\} \tag{2}$$

where $p_{U,j}^i$ denotes the j-th element of vector p_U^i. The greater threshold t means higher reliability requirement for the utilized unlabelled samples.

Merge the Class Certainty of Labelled Samples. Assuming that $L^i, i \in \{1, 2, \cdots, k\}$ is the original labelled sample set of class i. It's obvious that the corresponding class i certainty is 1 and the other class certainty is 0. Thus, the class i certainty vector p_L^i of L can be derived as,

$$p_{L,j}^i = \begin{cases} 1, & y_j = i \\ 0, & y_j \neq i \end{cases}, \; i \in \{1, 2, \cdots, k\}, \; j \in \{1, 2, \cdots, l\} \tag{3}$$

where $p^i_{L,j}$ denotes the j-th element of vector p^i_L and y_j denotes the label of x_j, respectively. By combining p^i_L and p^i_U, the class i certainty vector of training data $X = [L, U]$ is obtained:

$$p^i = [p^i_L, p^i_U],\ i \in \{1, 2, \cdots, k\} \tag{4}$$

2.3 New Semi-supervised LDA

In this section, we propose a novel semi-supervised LDA method by integrating class certainty into the scatter matrixes so that the samples are more separable in the projection subspace. At first, we define the within-class mean vector u_i and the total mean vector u,

$$\begin{aligned} u_i &= \frac{\sum_{j=1}^{N} p^i_j x_j}{\sum_{j=1}^{N} p^i_j} = X\left(p^i_j / \sum_{j=1}^{N} p^i_j\right) = X\widetilde{p^i} \\ u &= \frac{\sum_{i=1}^{K}\sum_{j=1}^{N} p^i_j x_j}{\sum_{i=1}^{K}\sum_{j=1}^{N} p^i_j} = X\left(\sum_{i=1}^{K} p^i / \sum_{i=1}^{K}\sum_{j=1}^{N} p^i_j\right) = X\widetilde{p} \end{aligned} \tag{5}$$

where p^i_j is the element of vector p^i.

Next, to obtain the "generalized Rayleigh quotient" of semi-supervised LDA, the new between-class scatter matrix S_b, within-class scatter matrix S_w and total-class scatter matrix matrixes S_t are defined as:

$$\begin{aligned} S_b &= \sum_{i=1}^{K} n_i (u_i - u)(u_i - u)^T \\ &= X[\sum_{i=1}^{K} m_i (\tilde{p}^i - \tilde{p})(\tilde{p}^i - \tilde{p})^T]X^T = X\tilde{S}_b X^T \end{aligned} \tag{6}$$

where $n_i = \sum_{j=1}^{N} p^i_j$.

$$\begin{aligned} S_w &= \sum_{i=1}^{k}\sum_{j=1}^{N} p^i_j (x_j - u_i)(x_j - u_i)^T \\ &= X[\sum_{i=1}^{k}\sum_{j=1}^{N} p^i_j (h_j - \tilde{p}^i)(h_j - \tilde{p}^i)^T]X^T = X\tilde{S}_w X^T \end{aligned} \tag{7}$$

$$S_t = \sum_{i=1}^{k}\sum_{j=1}^{N} p_j^i (x_j - u)(x_j - u)^T$$
$$= X[\sum_{i=1}^{k}\sum_{j=1}^{N} p_j^i (h_j - \tilde{p})(h_j - \tilde{p})^T]X^T = X\tilde{S}_t X^T \tag{8}$$

where $h_j \in R^{N\times 1}$ is expressed as:

$$h_{j,i} = \begin{cases} 1, & i = j \\ 0, & else \end{cases} \tag{9}$$

and $h_{j,i}$ denotes the element of h_j.

Since the new scatter matrixes have been proven to satisfy $S_t = S_b + S_w$, any two scatter matrixes can be utilized to construct the "generalized Rayleigh quotient". Generally, it is expressed in the following criterion,

$$\max_w \frac{w^T S_b w}{w^T S_w w} \tag{10}$$

where $w \in R^{d\times(k-1)}$ is the projection matrix. Then, w can be calculated by (11)

$$S_b w = \lambda S_w w \tag{11}$$

The closed-form solution of w related to k − 1 characteristic vectors of $S_w^{-1}S_b$.

2.4 Classification

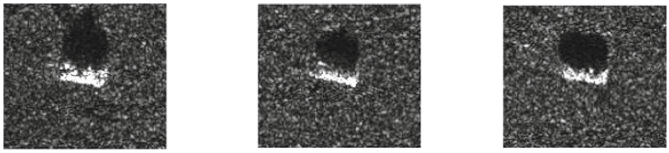
After the dimension reduction, the test data will be classified. There are several classifiers to be selected, such as SVM, random forest, nearest neighbor classifier (NNC) and so on. We adopts the NNC in this part because the training samples of the same class in the projection subspace are very close, which makes the mean vectors fully represent the characteristic information of every class. The mean vectors $\tilde{u}_i$ after dimension reduction can be expressed as

$$\tilde{u}_i = w^T u_i,\ i \in \{1, 2, \cdots, k\} \tag{12}$$

Then the class of test sample is determined by the nearest $\tilde{u}_i$.

3 Experiment

In this section, the performance of the proposed method is investigated on the Moving and Stationary Target Acquisition and Recognition (MSTAR) database. The discussion of CCS is performed initially to demonstrate the feasibility of CCS-related steps. Subsequently, the effectiveness of the proposed method is verified by comparing CCS with other semi-supervised algorithms. As shown in Fig. 3, we choose BMP2 (sn-c21), T72 (sn-132) and BTR70 (sn-c21) as the training data in the following experiments. Meanwhile, we select BMP2 (sn-c9566), T72 (sn-s7) and BTR70 (sn-c70) as the testing data. Table 1 lists the number of vehicles in the aforementioned dataset.



(a) T72 (b) BMP2 (c) BTR70

Fig. 3. The SAR images of three classes of vehicles.

Table 1. Types and quantities of training data and testing data.

	Training data			Testing data		
Type	T72	BMP2	BTR70	T72	BMP2	BTR70
Model	sn_132	sn_c21	sn_c71	sn_s7	sn_9566	sn_c70
Quantity	232	232	232	191	191	191

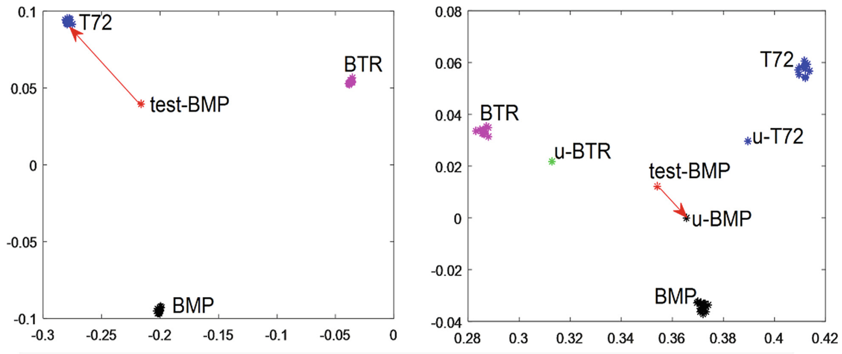
3.1 Discussion of CCS

To demonstrate the effectiveness of the semi-supervised LDA method, we compare it with the LDA method which only utilizes the labelled samples. We select 10% of the training data as the labelled samples and the remaining data as the unlabelled samples. As shown in Fig. 4, test-BMP represents a testing sample selected from the BMP vehicles. BMP, BTR, T72 denote the class mean vectors of the three types of vehicles obtained by the LDA method, and u-BMP, u-BTR, u-T72 denote the class mean vectors obtained by the semi-supervised LDA method. The direction of arrow represents the class judgment result of the test-BMP.

In Fig. 4(a), since test-BMP is closest to the mean vector of T72, the classification result is mistaken. Different from LDA, semi-supervised LDA can represent the truer feature distribution of samples by absorbing the characteristic information of unlabelled samples. As presented in Fig. 4(b), the test-BMP is obviously closest to the u-BMP and is correctly classified into BMP.

When obtaining the class certainty of unlabelled samples, we utilize the threshold processing to ensure the sufficient class reliability. Next, we discuss the impact of changing the threshold t on the performance of CCS. With the change of the percentage of labelled samples, the overall accuracy of different threshold is shown in Fig. 5.

When the percentage of labelled samples is small, the performance of CCS with $t = 0.2$ is highest, and the performance with $t = 0$ is the second highest. If the thresholds are relative big, which are set as 0.7 and 1, the classification performance of CCS is not good. The experimental result shows that when labelled samples are insufficient, compared with $t = 0$ which exploits all the class certainty information of unlabelled samples, setting t as a small value helps to improve the classification performance, which ensures the reliability of class certainty used in the semi-supervised LDA.



(a) LDA (b) semi-supervised LDA

Fig. 4. The effectiveness of the semi-supervised LDA method compared with the LDA method.

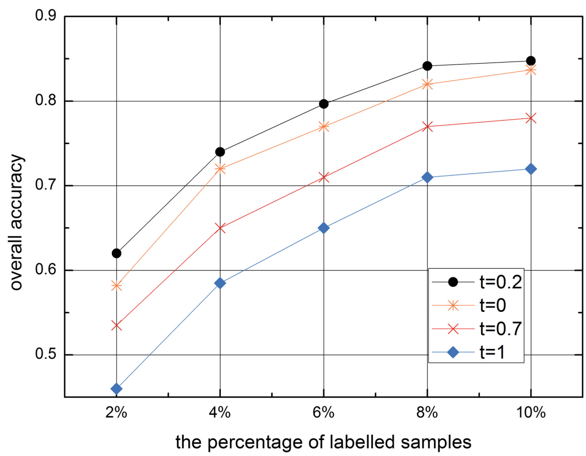


Fig. 5. Classification performance of CCS with different threshold setting.

3.2 Comparison with Other Semi-supervised Algorithms

In this section, we compare the performance of our method with that of the label propagation (LP) [14], progressive semi-supervised SVM with diversity (PS3VM-D) [12], constrained KMeans (C-KMeans) [15] and semi-supervised discriminant analysis

(SDA) [16]. As the percentage of labelled samples changes, the overall accuracy of different methods can be derived, as shown in Fig. 6.

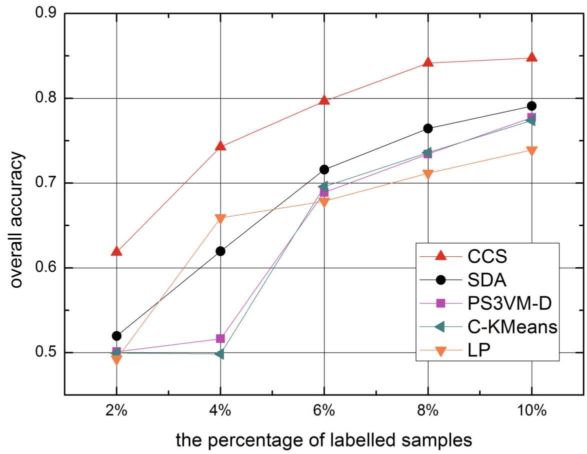


Fig. 6. Classification performance of CCS and the other four semi-supervised algorithms.

Obviously, the classification accuracy of CCS outperforms the other four semi-supervised algorithms by at least 8% when the labelled samples are insufficient. Generally, LP and PS3VM-D assign pseudo labels to unlabelled samples. However, the wrong pseudo labels will cause a bad influence on subsequent classifier training process. In terms of the C-KMeans, it can't make full use the spectrum information by adding constraints, which leads to little performance improvement. As for SDA, it focuses on maintaining the neighborhood relationship between samples, but has a high requirement of data distribution. Compared with the aforementioned four methods, CCS not only utilizes the class information of labelled samples, but also reliably absorbs the characteristic information of unlabelled samples through integrating the class certainty of samples into LDA, which makes the classification performance more stable and accurate.

4 Conclusion

To effectively solve the problem of RS images classification when labelled samples are insufficient, this paper proposes a novel semi-supervised classification method (CCS). There are three major findings:

(a) Based on the dimensional reduction of training samples and multi-class SVM based learning, the class certainty information is obtained and assigned to unlabelled samples for further processing.
(b) The pre-processed class certainty reassigns the weight for the unlabelled samples by normalizing and threshold processing.

(c) By combining class certainty, the proposed LDA can make full use of class information of labelled samples while characterizing reliably unlabelled samples.

From the experiment results, we observe that the CCS significantly improves the classification accuracy of RS images when the labelled samples are insufficient.

References

1. Zabalza, J., Ren, J., Zheng, J., Han, J.: Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging. IEEE Trans. Geosci. Remote Sens. **53**(8), 4418–4433 (2015)
2. Zhao, C., Li, X., Ren, J., Marshall, S.: Improved sparse representation using adaptive spatial support for effective target detection in hyperspectral imagery. Int. J. Remote Sens. **34**(24), 8669–8684 (2013)
3. Han, J.: Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. IEEE Trans. Geosci. Remote Sens. **53**(6), 3325–3337 (2015)
4. Yan, Y.: Unsupervised image saliency detection with Gestalt-laws guided optimization and visual attention based refinement. Pattern Recogn. **79**, 65–78 (2018)
5. Wang, Z.: A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos. Neurocomputing **287**, 68–83 (2018)
6. Bian, X., Zhang, T., Zhang, X.: Clustering-based extraction of near border data samples for remote sensing image classification. Cogn. Comput. **5**(1), 19–31 (2013)
7. Cao, F.: Sparse representation-based augmented multinomial logistic extreme learning machine with weighted composite features for spectral-spatial classification of hyperspectral images. IEEE Trans. Geosci. Remote Sens. **99**, 1–17 (2018)
8. Pasolli, E., Melgani, F., Tuia, D., Pacifici, F., Emery, W.J.: SVM active learning approach for image classification using spatial information. IEEE Trans. Geosci. Remote Sens. **52**(4), 2217–2233 (2014)
9. Blum, A., Chawla, S.: Learning from labeled and unlabeled data using graph mincuts. In: Eighteenth International Conference on Machine Learning, pp. 19–26. Morgan Kaufmann Publishers, USA (2001)
10. Blum, A.: Combining labeled and unlabeled data with co-training. In: Proceedings of the Eleventh Annual Conference on Computational Learning Theory, pp. 92–100 (2000)
11. Joachims, T.: Transductive inference for text classification using support vector machines. In: Sixteenth International Conference on Machine Learning, pp. 200–209. Morgan Kaufmann Publishers, Slovenia (1999)
12. Persello, C., Bruzzone, L.: Active and semisupervised learning for the classification of remote sensing images. IEEE Trans. Geosci. Remote Sens. **52**(11), 6937–6956 (2014)
13. Zhi-Hua, Z., Ming, L.: Tri-training: exploiting unlabeled data using three classifiers. IEEE Trans. Knowl. Data Eng. **17**(11), 1529–1541 (2005)
14. Wang, F., Zhang, C.: Label propagation through linear neighborhoods. IEEE Trans. Knowl. Data Eng. **20**(1), 55–67 (2007)
15. Wagstaff, K., Cardie, C., Rogers, S.: Constrained K-means clustering with background knowledge. In: Eighteenth International Conference on Machine Learning, pp. 577–584. Morgan Kaufmann Publishers, USA (2001)
16. Cai, D., He, X., Han, J.: Semi-supervised discriminant analysis. In: 11th International Conference on Computer Vision, pp. 1–7. IEEE, Brazil (2007)