CrossMark

THEORETICAL ADVANCES

Semi-supervised orthogonal discriminant projection for plant leaf classification

Shanwen Zhang[1] · Yingke Lei[2] · Chuanlei Zhang[3] · Yihua Hu[2]

Abstract Plant classification based on the leaf images is an important and tough task. For leaf classification problem, in this paper, a new weight measure is presented, and then a dimensional reduction algorithm, named semi-supervised orthogonal discriminant projection (SSODP), is proposed. SSODP makes full use of both the labeled and unlabeled data to construct the weight by incorporating the reliability information, the local neighborhood structure and the class information of the data. The experimental results on the two public plant leaf databases demonstrate that SSODP is more effective in terms of plant leaf classification rate.

Keywords Plant leaf classification · Dimensionality reduction · Orthogonal discriminant projection · Semi-supervised orthogonal discriminant projection

1 Introduction

Plant classification based on the leaf images is very important and necessary with respect to agricultural information, ecological protection and plant automatic classification systems. It is well known that how to validly extract classification features is central to the plant recognition based on leaf images. Currently, the widely used features for plant recognition based on leaf image can be divided into color, shape, and texture features [1–13]. However, these techniques have not fully considered the following characteristics of plant leaf image data. (1) Diversity: plant leaf images differ from each other in a thousand ways. On the one hand, the same kind of plant leaves appear in many different colors, shapes and sizes (as shown in Fig. 1a); on the other hand, different plants sometimes have a similar shape (as shown in Fig. 1b). (2) Nonlinearity: plant leaves vary with period, location, and illumination conditions, and make them distributed under some nonlinear manifold. These properties lead to the existing plant classifying methods difficult to achieve satisfactory recognition results.

In plant leaf recognition and classification tasks, a key process is to reduce the dimensionality of these high-dimensionality unorganized leaf image data. Effective dimensional reduction algorithm can help to avoid the "curse of dimensionality" problem in making decisions and analyzing data efficiently. Over the past several decades, many dimensional reduction methods have been proposed. These methods could be divided into three categories according to whether utilizing the class information of the input samples: unsupervised, supervised and semi-supervised. Supervised methods have been proven to be more suitable for classification tasks than those unsupervised ones. In classification problems, the label information can be used to guide the procedure of dimensionality reduction. Among the existing methods, LDA [14] is one of the most commonly used linear supervised methods and has been widely applied to many high-dimensional classification problems. It encodes discriminant information by finding directions that maximize the ratio of between-class scatter to within-class scatter. However, there are two major drawbacks with LDA: the small-sample-size (SSS)

✉ Yihua Hu
  eeibioinformatics@gmail.com

1 Department of Electronics and Information Engineering, Xijing University, Xi'an, China

2 Electronic Engineering Institute, Hefei, China

3 School of Computer Science and Information Engineering, Tianjin University of Science & Technology, Tianjin, China

Fig. 1 Sample images of plant leaves. **a** Large within-class difference, **b** large between-class similarity

(A) Large within-class difference

(B) Large between-class similarity
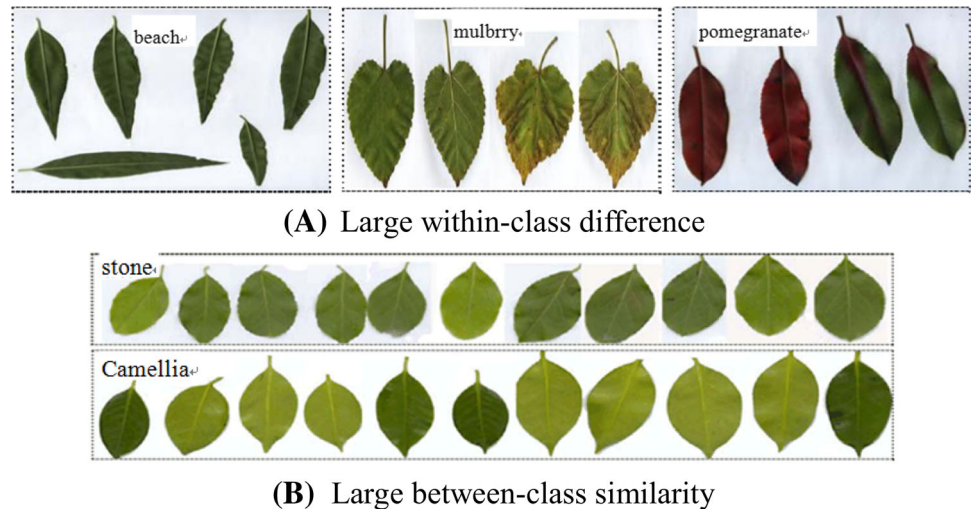
problem and the Gaussian distribution assumption. Marginal fisher analysis (MFA) [15] is a supervised manifold learning-based method, which without the prior information on data distributions, the inter-class margin can better characterize the separability of different classes than the inter-class variance as in LDA. However, a disadvantage of MFA is that there are many parameters to be set empirically. In contrast to MFA, only one K-nearest-neighbor criterion is introduced to define the neighborhood graph in locality sensitive discriminant analysis (LSDA) [16]. Orthogonal discriminant projection (ODP) is another supervised manifold learning method [17]. ODP follows the supervised techniques and the label information is taken to model the manifold. In ODP, the weight between any two points is defined by using both local information and non-local information to explore the intrinsic structure of original data and to enhance the recognition ability.

It is well known that it is much easier to collect the unlabeled data than the labeled samples. Recently, semi-supervised learning has emerged as a hot topic in pattern recognition area. The semi-supervised learning algorithms effectively utilize a limited number of labeled samples and a large amount of unlabeled samples for real-world applications. Cai et al. [18] proposed a semi-supervised discriminant analysis (SDA), in which the labeled data points are used to maximize the discriminating ability, while the unlabeled data points are used to maximize the locality preserving power. Based on LSDA, Zhao et al. [19] proposed a locality sensitive semi-supervised (LSSS) feature selection method. Like LSDA, LSSS does not effectively consider the global information of the data. Zhang et al. [20] proposed a semi-supervised locally discriminant projection (SSLDP). SSLDP is sensitive to noise points. In our paper, like many manifold learning-based dimensionality reduction methods, the noise points are the leaves with

Fig. 2 Two normal and four noisy samples of cherry blossom leaves

poor quality, which may be caused by period, location, illumination, or other factors. Figure 2 illustrates some samples from the cherry blossoms leaves, where the first and second are normal, the others are noise points, which are very different from the normal samples.

All the above-mentioned methods are summarized in the following Table 1.

Inspired by semi-supervised learning and ODP, we proposed a semi-supervised orthogonal discriminant projection (SSODP) method for plant leaf classification problem. In contrast to the original ODP algorithm, SSODP has the following desirable properties: (1) SSODP definitely utilizes the unlabeled sample information; (2) SSODP improves the weight construction by incorporating the reliability information.

The rest of the paper is organized as follows: ODP is briefly introduced in Sect. 2. SSODP is proposed in Sect. 3. Section 4 reports some experimental results based on two real plant leaf databases to demonstrate the effectiveness of the proposed method. Finally, conclusions and future work are drawn in the last section.

2 Orthogonal discriminant projection (ODP)

Assume there are n D-dimensional data points $X_1, X_2, \ldots X_n$, it is desirable to project these points into a linear subspace where the points with the same label are

Table 1 Summary of the dimensionality reduction methods

Name of methods	Disadvantages
Linear discriminant analysis (LDA)	The small-sample-size (SSS) problem and the Gaussian distribution assumption
Marginal fisher analysis (MFA)	There are many parameters needed to be set empirically
Locality sensitive discriminant analysis (LSDA)	The global information of data is not taken into account
Orthogonal discriminant projection (ODP)	It does not use the unlabeled sample information
Semi-supervised discriminant analysis (SDA)	The SSS problem and the Gaussian distribution assumption
Locality sensitive semi-supervised (LSSS)	It does not effectively consider the global information of the data
Semi-supervised locally discriminant projection (SSLDP)	It is sensitive to noise points

clustered closer and the points belonging to different classes are located farther. Let $Y_1, Y_2, \ldots, Y_n$ denote the projections of $X_1, X_2, \ldots, X_n$ in d-dimensional subspace, thus the projection can be expressed by $Y_i = A^T X_i$, where A is a linear transformation matrix. ODP seeks to find the linear transformation through the following steps.

Firstly, an adjacency graph $G = (V, E, H)$ is constructed using k nearest neighbor criterion, where G denotes the graph, V is the node set and E is the edge set. H is an adjacency matrix, whose elements are defined based on their both local information and class information:

$$H_{ij} = \begin{cases} E(X_i, X_j), & \text{If } X_j \in N(X_i) \text{ or } X_i \in N(X_j) \quad \text{and} \quad C_i = C_j \\ E(X_i, X_j) \cdot (1 - E(X_i, X_j)), & \text{If } X_j \in N(X_i) \text{ or } X_i \in N(X_j) \quad \text{and} \quad C_i \neq C_j, \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $E(X_i, X_j) = \exp[-d^2(X_i, X_j)/\beta^2]$, $d(X_i, X_j)$ denotes the Euclidean distance between points X_i and X_j, $N(X_i)$ is the k nearest neighborhoods of X_i, C_i is the label of X_i, and β is a parameter which is used as a regulator.

By introducing the weight matrix $H = \{H_{ij}\}$, the local scatter can be expressed to:

$$\begin{aligned} J_L(A) &= \frac{1}{2}\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}H_{ij}(Y_i - Y_j)^2 \\ &= \frac{1}{2}\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}H_{ij}(A^T X_i - A^T X_j)^T \\ &= \frac{1}{2}\frac{1}{n}\frac{1}{n}A^T\sum_{i=1}^{n}\sum_{j=1}^{n}H_{ij}(X_i - X_j)(X_i - X_j)^T A \\ &= A^T S_L A \end{aligned} \tag{2}$$

where $S_L = \frac{1}{2}\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}H_{ij}(X_i - X_j)(X_i - X_j)^T$.

After characterizing the local scatter, the non-local scatter can be characterized by the following expression,

$$\begin{aligned} J_N(A) &= \frac{1}{2}\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(1 - H_{ij})(Y_i - Y_j)^2 \\ &= \frac{1}{2}\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(Y_i - Y_j)^2 - \frac{1}{2}\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}H_{ij}(Y_i - Y_j)^2 \\ &= A^T S_T A - A^T S_L A, \end{aligned} \tag{3}$$

where $S_T = \frac{1}{2}\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - X_j)(X_i - X_j)^T$.

After the local scatter and the non-local scatter have been constructed, an optimization objective function is devised to maximize the difference between them with the constraint of $A^T A = I$. The objective function can be expressed as follows:

$$\arg\max_{A^T A = I} A^T((1 - \alpha)S_T - \alpha S_L)A, \tag{4}$$

where α is an adjustable parameter.

The projecting matrix A is composed of the d eigenvectors associated with the d top eigenvalues of the following eigenvalue equation,

$$((1 - \alpha)S_T - \alpha S_L')A = \lambda A. \tag{5}$$

Thus, the optimal linear features Y can be obtained by the following linear transformation:

$$Y = A^T X. \tag{6}$$

3 Semi-supervised orthogonal discriminant projection (SSODP)

Motivated by Eq. (1), we present a new weight between any two points as follows:

include at least two neighbors in each neighborhood The justification for using local neighborhoods is based on the argument that the weights should depend on the local geometry rather than the global structure of the whole data set.

$$W_{ij} = \begin{cases} \underset{i}{\text{Max}}(D_i) \cdot E(X_i, X_j) \cdot \left(1 + E(X_i, X_j)\right), & \text{If } X_j \in N(X_i) \text{ or } X_i \in N(X_j) \quad \text{and} \quad C_i = C_j, i \neq j \\ \underset{i}{\text{Min}}(D_i) \cdot E(X_i, X_j) \cdot \left(1 - E(X_i, X_j)\right), & \text{If } X_j \in N(X_i) \text{ or } X_i \in N(X_j) \quad \text{and} \quad C_i \neq C_j \\ \underset{X_j \in N(X_i) \text{ or } X_i \in N(X_j)}{\text{Mean}} [E(X_i, X_j)] \cdot E(X_i, X_j), & \text{If } X_j \in N(X_i) \text{ or } X_i \in N(X_j) \quad \text{and} \quad X_i \text{ or } X_j \text{ is unlabeled} \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

where $D_i = \frac{1}{k}\sum_{X_j \in N(X_i)} E(X_i, X_j)$ is the reliability of X_i $(i = 1, 2, \ldots, n)$, $\text{Max}_i(D_i)$ and $\text{Min}_i(D_i)$ is the maximum and minimum of $D_i(i = 1, 2, \ldots, n)$, respectively. $\text{Mean}_{X_j \in N(X_i) \text{ or } X_i \in N(X_j)} E(X_i, X_j) = \frac{1}{2k}\sum_{X_j \in N(X_i) \text{ or } X_i \in N(X_j)} E(X_i, X_j)$ is the weight mean of X_j in $N(X_i)$ or X_i in $N(X_j)$, and make $W_{ii} = 0$ to avoid the self-similarity of data points.

Figure 3 shows the typical plot of W_{ij} as a function of $d^2(X_i, X_j)/\beta^2$, where supposing $D_i = 1(i = 1, 2, \ldots, n)$, so $\text{Max}_i(D_i) = \text{Min}_i(D_i) = \text{Mean}(D_i) = 1$.

In Fig. 3, the curve A1 denotes X_j is among the k nearest neighborhoods of X_i or X_i is among the k nearest neighborhoods of X_j and they belong to the same class. The curve A2 denotes X_j is among the k nearest neighborhoods of X_i or X_i is among the k nearest neighborhoods of X_j and they do not belong to the same class. The curve A3 denotes X_j is among the k nearest neighborhoods of X_i or X_i is among the k nearest neighborhoods of X_j and X_i or X_j is unlabeled. The curve A4 denotes the other cases, including $W_{ij} = 0$ to avoid self-loops.

According to the Eq. (7), W_{ij} indicates that the neighboring points have more impact on the weight than the far-apart points. W_{ij} reflects not only the class information and reliability information of the data, but also the basic neighborhood relation between two points. This property demonstrates the discriminant similarity of the data. Since the Euclidean distance $d(X_i, X_j)$ is in the exponent, the parameter β is used to prevent $d^2(X_i, X_j)$ from increasing too fast when $d^2(X_i, X_j)$ is relatively large. The properties of W_{ij} can be summarized as follows:

1. By using the same neighborhood size for all points in a data set the neighborhood is just large enough to

2. In Eq. (7), D_i expresses the reliability information; $E(X_i, X_j)$ expresses local neighbor information; $1 + E(X_i, X_j)$ and $1 - E(X_i, X_j)$ express the discriminant information.

3. D_i of X_i is large if its k nearest neighbors are all closer to its vicinity, and small if they are far from its vicinity.

4. Since $1 < 1 + E(X_i, X_j) < 2$ and $0 < 1 - E(X_i, X_j) < 1$, when the Euclidean distance is equal, the within-class weight is larger than the between-class weight. This characteristic provides a certain chance to the points in the same class to have the larger weight than those in different classes, which is preferred for classification tasks.

5. With the decrease of the Euclidean distance, the between-class weight decreases toward 0. It means for the closer points from different classes that the larger $E(X_i, X_j)$, the smaller $1 - E(X_i, X_j)$ and the smaller value of weight.

With the new definition of the weight W_{ij}, the new semi-supervised dimensionality reduction algorithm is proposed as follows:

Due to the above properties of W_{ij}, we apply $W = [W_{ij}]$ to the original ODP for improving the discriminant ability of the dimensionality reduction methods. Given n D-dimensional data points $X = (X_1, X_2, \ldots, X_n)$, each column vector X_i represents a sample, we can construct an undirected full connected graph $G = (V, E, W)$, where each vertex represents a sample in X and the element W_{ij} defined as Eq. (7) in matrix W is the weight of edge between node X_i and X_j. W_{ij} can reflect the similarity between X_i and X_j. Considering a linear projection $A \in R^{n \times d}$ that maps each point X_i in the original space to a lower dimensionality vector $Y_i \in R^d$, $Y_i = A^T X_i$, the local scatter J'_L and non-local scatter J'_N can be defined as follows:

Fig. 3 Typical plot of W_{ij} as a function of $d^2(X_i, X_j)/\beta^2$
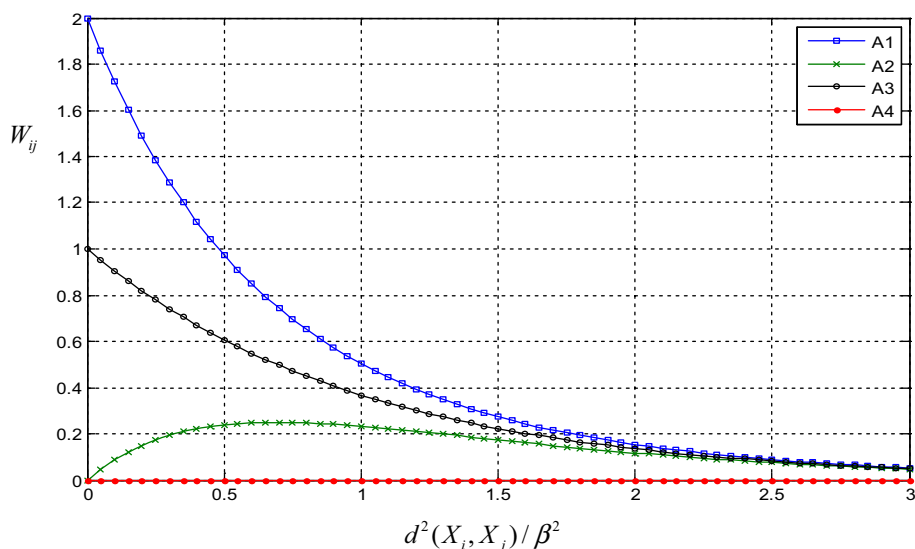
$$J'_L(A) = \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}(Y_i - Y_j)^2 \qquad (8)$$

$$J'_N(A) = \sum_{i=1}^{n} \sum_{j=1}^{n} (1 - W_{ij})(Y_i - Y_j)^2 \qquad (9)$$

By simple algebraic formulation, Eq. (8) can be formulated as:

$$\frac{1}{2}J'_L(A) = \frac{1}{2}\sum_i \sum_j W_{ij}(A^T X_i - A^T X_j)^2$$
$$= \sum_{i=1}^{n} \sum_{i=1}^{n} A^T X_i D_{ii} X_i^T A - \sum_{i=1}^{n} \sum_{j=1}^{n} A^T X_i W_{ij} X_j^T A$$
$$= A^T X S'_L X^T A \qquad (10)$$

where $S'_L = D - W$ is the graph Laplacian of the within-class graph and $W = \{W_{ij}\}$, D is a diagonal matrix with entries $D_{ii} = \sum_j W_{ij}$.

Similarly, Eq. (9) can be formulated as:

$$\frac{1}{2}J'_N(A) = \frac{1}{2}\sum_i \sum_j (1 - W_{ij})(A^T X_i - A^T X_j)^2$$
$$= A^T X(E_{ii} - F_{ij})X^T A \qquad (11)$$
$$= A^T X S'_N X^T A,$$

where $S'_N = E - F$ is the graph Laplacian of the between-class graph and, $F = \{1 - W_{ij}\}$, E is a diagonal matrix with entries $E = \sum_j (1 - W_{ij})$.

After the local scatter J'_L and the non-local scatter J'_N have been constructed, the objective function is written as:

$$\arg\max_A \frac{tr(A^T X S'_N X^T A)}{tr(A^T X S'_L X^T A)} \qquad (12)$$
$$s.t. A^T A = I.$$

The Lagrangian function of the problem Eq. (10) is

$$\arg\max_A \frac{tr(A^T X S'_N X^T A)}{tr(A^T X S'_L X^T A)} - \lambda(A^T A - I) = 0. \qquad (13)$$

In order to get the dth orthogonal basis vector, we maximize the following objective function:

$$\arg\max_A \frac{tr(A^T X S'_N X^T A)}{tr(A^T X S'_L X^T A)} \qquad (14)$$

with the constraints $a_d^T a_1 = a_d^T a_2 = \cdots = a_d^T a_{d-1} = 0$, where d is the output reduction dimensionality.

To solve the above optimization problem, we use the Lagrangian multiplier:

$$f = a_d^T (X S'_N X^T - \delta X S'_L X^T)a_d - \lambda_1 a_d^T a_1 - \lambda_2 a_d^T a_2 - \cdots$$
$$- \lambda_{d-1} a_d^T a_{d-1} \qquad (15)$$

where δ is a number to be determined. Set the partial derivative of f with respect to a_d to zero and obtain

$$2(X S'_N X^T - \delta X S'_L X^T)a_d - \lambda_1 a_1 - \lambda_2 a_2 - \cdots - \lambda_{d-1} a_{d-1} = 0 \qquad (16)$$

Multiplying the left side of Eq. (14) by a_d^T, we obtain

$$a_d^T (X S'_N X^T - \delta X S'_L X^T)a_d = 0 \Rightarrow \delta = \frac{tr(A^T X S'_N X^T A)}{tr(A^T X S'_L X^T A)}. \qquad (17)$$

According to the paper [21, 22], the orthogonal iterative algorithm of solving the orthogonal vectors of SSODP is summarized as follows:

1. Initialize A as an arbitrary column-orthogonal matrix.
2. Compute the trace ratio value $\delta = \frac{tr(A^T X S'_N X^T A)}{tr(A^T X S'_L X^T A)}$.
3. Form the trace difference problem as $A' = \max_{A^T A = I} tr(A^T(X S'_N X^T - \delta X S'_L X^T)A)$.
4. Establish A' by the d eigenvectors of $X S'_N X^T - \delta X S'_L X^T$ corresponding to the d largest eigenvalues.
5. Update A by A'.
6. Iteratively perform steps (2)–(5) until convergence.
7. Output the final linear transformation matrix.

From the above iterative processes, we obtain the mapping matrix $A = (a_1, a_2, \ldots, a_d)$. The low-dimensional representation in the SSODP subspace is defined as follows:

$$X \rightarrow Y = A^T X \qquad (18)$$

4 Experiments

In this section, we apply the proposed method to plant leaf classification in two public leaf databases. One is ICL-Plant Leaf database (http://www.intelengine.cn/source.htm), which contains more than 30,000 leaf images of 362 plant species. All leaf images were captured in different seasons, and at different locations and natural illuminations. Another is Swedish leaf dataset. The Swedish leaf dataset [23] contains 1125 images of leaves uniformly distributed in 15 species. The Swedish leaf dataset is very challenging because of its high inter-species similarity.

In the two databases, the original leaf images contain footstalks or petioles, which are not suitable for robust leaf recognition, since the length and orientation of those footstalks may heavily depend on the collection process. So we crop and normalize (in scale and orientation) and resize them to 32×32 pixels by histogram equilibrium with 255 gray levels per pixel and with the white background. All leaf images are represented as points in vector space. A typical image of size 32×32 describes a point in 1024-dimensional vector space.

In the two databases, although all leaf images are provided with class labels, for each experiment, we randomly draw several samples from every database as a training set, among them, a random subset is used with labels as the labeled subset, and the remained are used as the unlabeled subset.

To verify the effectiveness of the proposed method, we compared it with RLDA [24], ODP [17], SDA [18], and

SSLDP [20]. ODP only makes use of the labeled data points. However, SDA, SSDR and our proposed method make use of both labeled and unlabeled data points. The training set is used to learn the low-dimensional subspace with the projection matrix. The testing set is used to obtain the final recognition rate. In every experiment, firstly, we randomly select an amount of images from all selected images as training set and the rest as testing set; secondly, we execute the dimensionality reduction method on the training dataset and obtain the mapping matrix; then the testing set is projected into the low-dimensional feature subspace; finally, the plant leaf images of the testing set are recognized by a suitable classifier. The parameters k and β are determined empirically. In the following experiments, the 1-nearest-neighbor (1-NN) classifier with Euclidean distance is adopted to classify plant leaf images. In these methods, PCA is used for data preprocessing (dimension reduction) in order to avoid the singular matrix problem and improve the computational efficiency.

4.1 Experiments on ICL-plant leaf database

In experiments, we select 500 leaf images from 20 kinds of plants and each plant has 25 leaf images. Figure 4 shows 20 representative leaf images of 20 plant species.

The classifying experiments are conducted to evaluate the effectiveness of the proposed method and to examine the effectiveness of the training number and the number of labeled points in training set on the performance. After all images preprocessed, from each kind of plant, we randomly select m leaf images as the training set, the rest $(25 - m)$ is regarded as the test set, among them, $l(l = 1, 2, 3, 4, 5, 6)$ labeled leaf images and $m - l$ unlabeled leaf images. For each given m and l, a trial is independently performed 50 times. The maximal average result of every experiment is recorded. The average recognition rate means the performance of the classifier with the best set of fitted parameters. The final result is the average correct recognition rate and standard derivation over 50 random
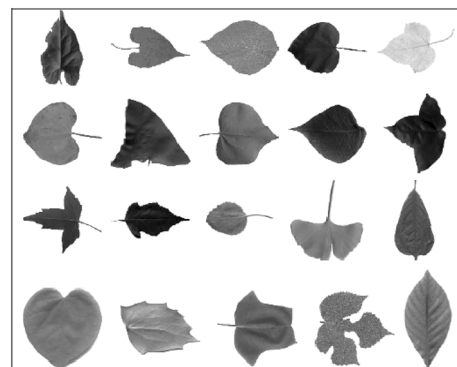

Fig. 4 The representative leaf images of 20 plant species

Table 2 Average correct recognition rates and standard derivation comparing with LPP, ODP, SDA, SSDR, SSLDP on the ICL-plant leaf database, where $k = 15$ and $\beta = 200$

Method	$m = 10$ $l = 1$	$m = 10$ $l = 2$	$m = 10$ $l = 3$	$m = 10$ $l = 4$	$m = 10$ $l = 5$	$m = 10$ $l = 6$
RLD2	0.2338	0.6379	0.7104	0.7843	0.8586	0.8764
	±0.021	±0.051	±0.042	±0.053	±0.038	±0.045
ODP	0.2174	0.5882	0.7372	0.7923	0.8712	0.8947
	±0.058	±0.024	±0.048	±0.048	±0.042	±0.032
SD2	0.5815	0.7145	0.7859	0.8243	0.9423	0.9223
	±0.053	±0.052	±0.046	±0.032	±0.042	±0.034
SSLDP	0.6712	0.7529	0.8584	0.9183	0.9385	0.9582
	±0.026	±0.039	±0.037	±0.023	±0.023	±0.024
SSODP	0.7528	0.7744	0.8963	0.9318	0.9638	0.9714
	±0.047	±0.022	±0.036	±0.023	±0.035	±0.031

Fig. 5 Plant leaf recognition rates versus the reduction dimensionality, where $m = 10$ and $l = 4$
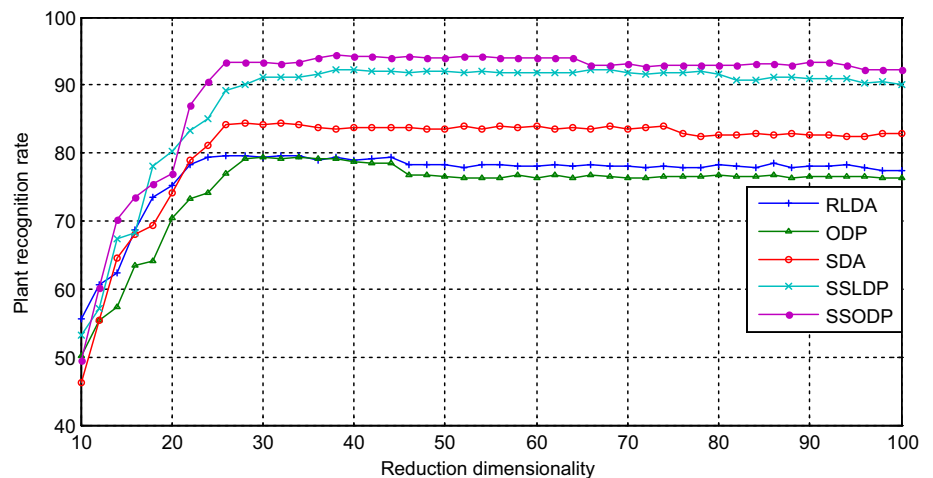
splits. We find that the proposed method is not sensitive to β when it is enough large, so we set $\beta = 200$. The other parameters k and d are decided by the maximal correct recognition results. Table 2 shows the performance of RLDA, ODP, SDA, SSLDP, and the proposed algorithm. As can be seen, our algorithm achieved 97.12 % recognition rate when six leaf labeled images per class were used for training, which is the best out of the six algorithms. We test the influence of selecting different dimensions in the reduced subspace on the recognition rate, as shown in Fig. 5, where $m = 10$ and $l = 4$. In general, the performance of all these methods varies with the number of dimensions. At the beginning, the recognition rates increase greatly as d increases. However, more dimensions do not lead to higher recognition rate after these methods attain the best results.

4.2 Experiments on Swedish leaf database

The Swedish leaf data set comes from a leaf classification project at Linköping University and the Swedish Museum of Natural History [23]. The data set contains the isolated
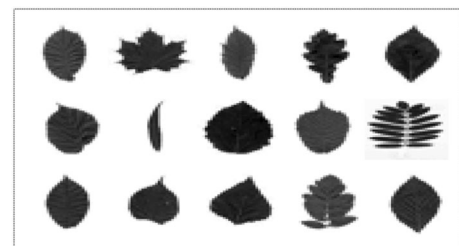

Fig. 6 The representative images of 15 different species in Swedish leaf database, one image per species. Note that some species are quite similar, e.g., the first, third, and ninth images

leaves from 15 different Swedish tree species, with 75 leaves per species. Figure 6 shows the representative leaf images of 15 different Swedish trees.

The experimental design is the same as before. In every experiment, this procedure is also repeated 50 times and we obtain 50 maximal results for the proposed algorithm and each of the compared methods. The average recognition rates and standard derivations are recorded, as seen in Table 3. From Table 3, we can see three main points. First, OLDSE-ISSODP outperforms the other algorithms among

Table 3 Average recognition correct rates and standard derivation comparing with LPP, ODP, SDA, SSDR, SSLDP on the Swedish leaf database, where $k = 10$ and $\beta = 200$

Method	$m = 10$ $l = 1$	$m = 10$ $l = 2$	$m = 10$ $l = 3$	$m = 10$ $l = 4$	$m = 10$ $l = 5$	$m = 10$ $l = 6$
RLDA	0.651	0.702	0.712	0.793	0.841	0.905
	±0.035	±0.041	±0.043	±0.045	±0.038	±0.035
ODP	0.614	0.724	0.732	0.796	0.854	0.912
	±0.037	±0.042	±0.038	±0.038	±0.031	±0.042
SDA	0.705	0.756	0.792	0.843	0.942	0.964
	±0.032	±0.047	±0.036	±0.034	±0.041	±0.034
SSLDP	0.747	0.791	0.912	0.948	0.963	0.978
	±0.036	±0.029	±0.037	±0.033	±0.037	±0.042
SSODP	0.783	0.802	0.922	0.955	0.987	0.988
	±0.037	±0.038	±0.037	±0.033	±0.025	±0.036

all the cases. For example, when the ten leaf images per species are used for training, SSODP leads RLDA, ODP, SDA, and SSLDP up to 1.12, 4.05, 3.11, and 0.69 %, respectively. Second, SDA, SSLDP, and SSODP perform better than RLDA and ODP. The reason is that the semi-supervised methods make use of all the training images (including labeled and unlabeled images).

5 Discussion

From Tables 2, 3 and Fig. 5, we can see that the semi-supervised methods outperform ODP. The reason is that the semi-supervised methods make use of all the training images (including labeled and unlabeled images), while ODP only use the labeled data points. It is worth pointing out that the recognition rates in the optimizing procedure mainly depend on the ways of splitting the given training data set and the number of the labeled points. From Tables 2 and 3, as more leaf images and the labeled data points split from the given data set are used for training, the recognition rates increase for all the compared algorithms. This demonstrates that all the algorithms are capable of learning from these data points. Particularly, when $l = 1$, the recognition rates of SSLDP and SSODP are more than 70 %, so the intrinsic image manifold can still be estimated even with a single labeled leaf image per class.

From Tables 2, 3 and Fig. 5, we can see our method outperforms the other methods. A possible explanation is that our method is more robust. In fact, there are some irregular-shaped leaf images in the two plant leaf data-bases, which were taken from the same tree or from the same leaf in different sessions, so the similarity (or weight) of intra-leaf may be smaller than that of inter-leaf, as seen Fig. 1. As a result, nearest neighborhood points may belong to the different classes. Moreover, the leaf images were taken during two different sessions, so that the appearance of the same leaf image may look different,

resulting in different distributions of the data points in the two sessions. The nearest neighbors of a point are more likely to belong to different classes. If points within the same nearest neighborhood are treated from the same class as ODP or SDA does, the recognition accuracy may be seriously affected. In our proposed method, we use $\text{Max}_i(D_i)$, $\text{Min}_i(D_i)$, and $\text{Mean}_{X_j \in N(X_i) \text{ or } X_i \in N(X_j)}$ to weaken the impact of the irregular-shaped or distorted leaves. In order to explain this conclusion, the experiments include several irregular-shaped leaf images in the training set and test set. To test whether our method is sensitive to such irregular-shaped leaf images, we remove them from the training set and conduct the experiments. The results indicate that the recognition rates are slightly affected by this kind of phenomenon. The reason might lie in the robust path-based similarity measurement [25]. In [25], the similarity expression $\alpha_i = \sum_{X_j \in N(X_i)} E(X_i, X_j)$ is adopted to improve the robustness of the algorithm to noise points.

6 Conclusions and future works

Semi-supervised manifold learning-based methods have aroused a great deal of interest in dimensionality reduction. In this paper, a semi-supervised dimensionality reduction algorithm named semi-supervised orthogonal discriminant projection (SSODP) was proposed and applied successfully to leaf classification. SSODP utilizes all labeled and unlabeled samples to construct the weight between any pairwise points. The experiment results on leaf images demonstrated that SSODP is effective and feasible for leaf classification.

In SSODP, the common assumption is that we only have a limited number of labeled training samples. Developing an adaptive algorithm to automatically adjust the parameters k and β is our future work. In addition, in this paper, we focus on developing a semi-supervised manifold learning technique for leaf classification but do not

systematically analyze the stability and robustness of the SSODP. This is a problem deserving further investigation.

References

1. Venters CC, Cooper DM (2000) A review of content-based image retrieval systems. Technical report, Manchester Visualization Centre, Manchester Computing, University of Manchester, Manchester, UK

2. Smeulders AWM, Worring M, Santini S et al (2000) Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell 22:1349–1380

3. Tico M, Haverinen T, Kuosmanen P (2000) A method of color histogram creation for image retrieval. In: Proceedings of the nordic signal processing symposium (NORSIG'2000), Kolmarden, Sweden, pp 157–160

4. El-ghazal A, Basir OA, Belkasim S (2007) Shape-based image retrieval using pair-wise candidate co-ranking. Lect Notes Comput Sci 4633:650–661

5. Park JS, Kim T-Y (2004) Shape-based image retrieval using invariant features. Advances in multimedia information processing-PCM, Berlin/Heidelberg. Lect Notes Comput Sci 2004:146–153

6. Abbasi S, Mokhtarian F, Kittler J (1999) Curvature scale space image in shape similarity retrieval. Multimed Syst 7:467–476

7. Ma W, Manjunath B (1996) Texture features and learning similarity. In: Proceedings of the conference on computer vision and pattern recognition (CVPR), pp 425–430

8. Han J, Ma K-K (2007) Rotation-invariant and scale-invariant gabor features for texture image retrieval. Image Vis Comput 25:1474–1481

9. Zhang D, Aylwin Wong MI, Lu G (2000) Content based image retrieval using gabor texture features. In: Proceedings of 1st IEEE Pacific Rim Conference on Multimedia (PCM00), pp 392–395

10. Manjunath B, Ma W (1996) Texture features for browsing and retrieval of image data. IEEE Trans Pattern Anal Mach Intell 18:837–842

11. Liu C, Wechsler H (2001) A Gabor feature classifier for face recognition. ICCV 2:270–275

12. Agarwal G, Belhumeur P, Feiner S et al (2006) First steps toward an electronic field guide for plants. Taxon 55(3):597–610

13. Kumar N, Belhumeur PN, Biswas A et al (2012) Leafsnap: a computer vision system for automatic plant species identification. Proc ECCV 2012:502–516

14. Belhumeour PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720

15. Yan SC, Xu D, Zhao BY et al (2007) Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans Pattern Anal Mach Intell 29(1):40–51

16. Cai D, He X, Zhou K, Han J, Bao H (2007) Locality sensitive discriminant analysis. IJCAI'07. In: Proceedings of the 20th international joint conference on Artificial intelligence, pp 708–713

17. Li B, Wang C, Huang DS (2009) Supervised feature extraction based on orthogonal discriminant projection. Neurocomputing 73:191–196

18. Cai D, He X, Han JW (2000) Semi-supervised discriminant analysis. IEEE 11th international conference on computer vision, pp 1–7

19. Zhao J, Lu K, He X (2008) Locality sensitive semi-supervised feature selection. Neurocomputing 71:1842–1849

20. Zhang SW, Lei YK, Wu YH (2011) Semi-supervised locally discriminant projection for classification and recognition. Knowl Based Syst 24:341–346

21. Zhu L, Zhu S (2007) Face recognition based on orthogonal discriminant locality preserving projections. Neurocomputing 70:1543–1546

22. Hu HF (2008) Orthogonal neighborhood preserving discriminant analysis for face recognition. Pattern Recogn 41:2045–2054

23. Söderkvist, O (2001) Computer vision classification of leaves from Swedish Trees. Master Thesis, Linköping Univ., Linköping

24. Friedman JH (1989) Regularized discriminant analysis. J Am Statist Assoc 84:165–175

25. Chang H, Yeung DY (2008) Robust path-based spectral clustering. Pattern Recogn 41:191–203