



Fuzzy weighted sparse reconstruction error-steered semi-supervised learning for face recognition

Li Liu¹ · Siqi Chen¹ · Xiuxiu Chen¹ · Tianshi Wang¹ · Long Zhang¹

Published online: 21 September 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Since the number of labeled data is limited in the semi-supervised learning settings, we propose a fuzzy weighted sparse reconstruction error-steered semi-supervised learning method for face recognition. The fuzzy membership functions are introduced to the reconstruction error calculation for the unlabeled data. A weight function is utilized to capture the locality property of data when learning the sparse coefficients. The fuzzy weighted sparse reconstruction error-steered semi-supervised learning not only inherits the advantages of sparse representation classification techniques and neighborhood methods, but also steers the reconstruction errors of unlabeled data. Experimental studies on well-known face image datasets demonstrate that the proposed method outperforms the comparative approaches.

Keywords Membership function · Sparse representation · Fuzzy

1 Introduction

In many real applications, labeling data is time-consuming and expensive, thus resulting in very few labeled data presented for data classification or clustering tasks. Since unlabeled data can be easily obtained in many problems, how to utilize both a small set of labeled data and a large amount of unlabeled data to classify the unlabeled data attracts considerable attentions from machine learning and pattern recognition communities. This learning schema is called semi-supervised learning (SSL), and its objective is accurately assigning class labels to the unlabeled data by incorporating the supervised information of the limited labeled data and their relationships to the unlabeled ones. Many efforts have been made on graph-based models in recent years since the intrinsic manifold structure of both the labeled and unlabeled data can be effectively considered. Graph-based semi-supervised learning (GSSL) uses a graph to represent the relational patterns among data, and a graph includes sets of vertices and edges connecting pairs of vertices. Each vertex denotes a datum, and each edge is assigned a weight to measure the similarity between a pair of data. Labels are

assigned to unlabeled data based on the assumptions that data connected by large weights should have similar label. Both the labeled and unlabeled data are incorporated in the graph to represent the adjacency relations among them. GSSL approaches meet the manifold assumption, and can well-approximate the intrinsically low-dimensional manifold data structure.

The success of GSSL mainly depends on the definition of similarity measure between pairs of data. The most commonly adopted similarity measure approaches are built on the basis of Gaussian kernel functions in Euclidean space, which solely depend on the Euclidean distance between two data and a kernel parameter [1]. The K-nearest neighbor (KNN) approach only assigns nonzero weights to edges connecting a datum and its K-nearest neighbors, but the relations among data include not only the neighborhood, but also the others, such as linear neighborhood embedding [2] and sparse representation [3], thus some more sophisticated methods are proposed due to the new relations for similarity measure [4, 5]. Wang et al. [5] formulated data similarities in a linear space by assuming that each datum could be approximately and linearly reconstructed by its neighbors, and Cheng et al. [4] applied sparse representation techniques to formulate a sparse similarity measure based on the assumption that each datum can be sparsely and linearly reconstructed by the elements of a dictionary. Compared with linear neighborhood propagation (LNP) [5],

✉ Li Liu
liuli_790209@163.com

¹ School of Information Science and Engineering, Shandong Normal University, Jinan 250014, Shandong Province, China

the similarity induced by sparsity [4, 6] measures the pair-wise similarities using sparse overcomplete code coefficients and does not need to predetermine the number of the nearest neighbors.

In order to boost the performance of GSSL, kernel tricks [7, 8] have been introduced to the formulation of data similarity measure. Some approaches integrate multiple similarity measures, multiple kernel functions or both of them into one model, so as to accurately mine the intrinsic data structure as much as possible [9–11]. The similarity measures focus on the aspects of data structure, and the kernel functions are utilized to project the data into different feature spaces, so the integration of these measures and functions inherits their advantages. Although existing graph-based methods boost the classification performance, they are still not sufficient to exhibit multiple relationships between pairs of data.

Since sparse representation (SR) techniques avoid the difficulty of determining the parameters used in KNN and ξ ball methods [2], some works have been done on this direction for classification or image recognition problems. Cheng et al. [4] first introduced a sparsity-induced similarity measure for label propagation, and redefined the similarity measure by reconstructing each data point as a l_1 -norm sparse linear combination of the rest ones. Since the sparse coefficients depict the spatial distribution of data and reflect the neighborhood relationship, they can be utilized to measure the similarities between pairs of data. Zhao et al. [12] proposed a forward and backward sparse representation method to obtain the degree of confidence with a specific class label for unlabeled data, and utilized co-training to build a novel semi-supervised learning approach to boost the image annotation performance. Based on the assumption that both the label space and the data space share the same sparse relationship and sparse reconstruction coefficients, Zhang et al. [13] utilized SR to learn the sparse coefficients for each datum in the original data space, and then used the sparse coefficients in the linear reconstruction of labels in the label space. The labels of unlabeled data can be obtained by minimizing the sum of label reconstruction errors of all data. Zhuang et al. [14] proposed a method to obtain a nonnegative low rank and sparse matrix, and used it for calculating the data similarity. The proposed similarity measure captures both the global and local data structure. In order to accurately capture the data structure, Peng et al. [15] described an enhanced low-rank data representation approach, and Fan et al. [16] introduced a sparse regularization penalty term to the manifold regularization by replacing the adjacency graphs of data, and obtained good performance for semi-supervised learning. A nuclear norm regularization was also proposed to simultaneously handle the pair-wise constraints among data and the sparse labeled data issues [17]. Kobayashi [9] described a similarity measure from the probabilistic

viewpoint, which could capture the sparse and KNN-like properties of data. Since the labeled data are quite limited in SSL, in order to obtain more supervised information for label propagation, Zhang et al. [18] utilized the propagated soft labels of both the labeled and unlabeled data to enrich the data relationship, and reported good performance for GSSL. Wang et al. [19] collected the positive and negative unlabeled samples, and used the labeled and unlabeled training samples to select the weak classifiers. Pei et al. [20] integrated the classical label propagation optimization function and the sparse coefficient learning lasso function into one model to simultaneously optimize the sparse coefficients and the labels of the unlabeled data.

Almost all the above approaches define novel weights for graph edges by incorporating SR techniques, and make label propagation among data in the label space based on the obtained weight matrix. Although they take into account the global and local data structure, the sparse reconstruction errors of data are ignored, which is also essential for sparse representation-based semi-supervised learning. Song et al. [21] proposed a fuzzy adaptive way to adapt dictionaries in order to achieve the fuzzy sparse signal representations. The update of the dictionary columns was combined with an update of the sparse representations by embedding a new mechanism of fuzzy set.

A good semi-supervised classification method should not only fully capture both the true intrinsic local and global structures of data, but also be steered by the classification task itself through incorporating the supervised information of the labeled data. Different from GSSL, which aims to find good similarity measure among data by neighborhood methods, sparse relationships or integration of label propagation and sparse constraints, we propose a novel approach which inherits the advantage of sparse representation techniques and neighborhood methods. This approach introduces fuzzy memberships into the calculation of reconstruction errors for unlabeled data, and steers unlabeled data classification based on the fuzzy memberships.

In order to clarify the contributions of our work, we briefly summarize the novel characteristics of the proposed approach as follows:

1. We introduce fuzzy memberships to the reconstruction error calculation and point out that the sparse representation classification implicitly takes into account all the labeled data in calculating the data reconstruction error associated with each class.
2. The proposed approach defines membership functions based on the fuzzy reconstruction errors of the unlabeled data, and alternately updates the fuzzy memberships until a termination condition is satisfied. Classification of the unlabeled data is based on the finally obtained fuzzy memberships. Available techniques can be uti-

lized to solve the optimization problems of the proposed method, and no new solution technique is needed.

This paper is organized as follows. Section 2 briefly introduces GSSL and the sparse learning for image recognition. In Sect. 3, we propose the fuzzy membership learning and updating approach with the label information. Section 4 reports the experimental results and the discussions about these results. Finally, the paper is concluded in Sect. 5.

2 Preliminaries

2.1 GSSL methods

GSSL approaches make use of the manifold structure embedded in the labeled and unlabeled data for realizing label information propagation over the graph. We give a brief review in the following.

A graph $G = (V, E)$ represents the relational patterns among data, where V and E are the set of vertices and edges connecting pairs of vertices. Given a semi-supervised c class classification task, we use both $X_l = \{x_1, x_2, \dots, x_{n_l}\}$ and $X_u = \{x_1, x_2, \dots, x_{n_u}\}$ to denote the small size labeled dataset and the unlabeled dataset, respectively, with $x_i \in R^q$, where q is the data dimensionality. Let $n = n_l + n_u$ be the size of $X (= X_l \cup X_u)$ with n_l labeled and n_u unlabeled ones, and both $F_l = \{f_1, f_2, \dots, f_{n_l}\}$ and $F_u = \{f_{n_l+1}, f_{n_l+2}, \dots, f_n\}$ be the label set of the labeled data and the unlabeled data, respectively, where f_i denotes the label vector of x_i . If x_i belongs to the k th class, then the k th entry of f_i equals 1, and all the other entries equal 0. A $n \times n$ symmetric weight matrix $S = [s_{ij}]$ is defined, where s_{ij} is a weight assigned to the edge connecting x_i and x_j . The label propagation is formulated as minimizing the following the energy function [22], and the labels of the unlabeled data can be obtained by solving it.

$$E = \frac{1}{2} \sum_{i,j} s_{ij} \|f_i - f_j\|^2 = \mathbf{FLF}^T \quad (i, j \in \{1, 2, \dots, n\}) \quad (1)$$

where \mathbf{F} is a label matrix with each sample label vector as one column, $\mathbf{L} = \mathbf{D} - \mathbf{S}$, and \mathbf{D} is a diagonal matrix with $d_{ii} = \sum_{j=1}^n s_{ij}$.

2.2 Sparse learning

The sparsity of the representation coefficients and the size of the dictionary play essential roles for the success of classification. But in many real applications, we have limited number of labeled data together with a large amount of unlabeled data, thus how to fully utilize the relational pattern together with the local and global structure of the data becomes an obstacle to improve the classification performance.

Since the overcomplete dictionary of SR depicts the distribution of the entire data with the sparse coefficients encoding the sparse neighborhood relationship of data, both of them should be fully utilized for SSL. In order to satisfy the manifold and cluster assumption, we multiply the sparse coefficients with weights which characterize the local relationship of data.

Wright et al. [23] used the sparse representation techniques for face image reconstruction and classification. Given a training dataset $X_1 = \{X_{11} \cup X_{12} \cup \dots \cup X_{1c}\} \in R^{q \times n_1}$, where c is the number of classes, X_{li} is the subset of X_l with all n_{li} data belonging to class i , and $n_l = \sum_{i=1}^c n_{li}$ is the number of training data in all classes. For a query datum y , the following two steps are conducted to identify its label:

1. Obtaining the sparse reconstruction coding coefficient vector by solving the following l_1 -norm minimization problem:

$$\min \|y - \mathbf{X}_l \alpha\|^2 + \gamma \|\alpha\|_1 \quad (2)$$

where γ is a constant that controls the sparsity of α , and \mathbf{X}_l denotes the matrix of the labeled data with each datum as a column.

2. After α has been obtained, calculating the reconstruction error associated with each class, and performing classification as

$$\text{label}(y) = \arg \min_i \|y - \mathbf{X}_{li} \alpha_i\|^2 \quad (i \in \{1, 2, \dots, c\}) \quad (3)$$

where α_i is the coding coefficients associated with \mathbf{X}_{li} , and \mathbf{X}_{li} denotes the matrix of the i th class labeled data with each datum as a column.

Equation (3) shows that Wright et al. [23] directly utilize the class related atoms of the overcomplete dictionary for calculating the reconstruction errors.

3 Learning the fuzzy memberships for unlabeled data

In order to learn the memberships associated with each class for the unlabeled data, we first initialize the memberships of all data based on the prior labeled information, and introduce a weight matrix to the sparse coefficient learning. After that, for each unlabeled datum, we calculate its reconstruction error associated with each class, and initialize its memberships based on the reconstruction errors. The memberships are introduced into the reconstruction error calculation in the following steps, and are updated based on the newly obtained reconstruction errors. This process is repeated until a termination criterion is reached.

3.1 Fuzzy membership initialization

Let $C = \{1, 2, \dots, c\}$ be the label set, for each labeled datum $x_i (i \in \{1, \dots, n_l\})$ and $\forall j \in C$; if x_i belongs to class p , we initialize its memberships as

$$m_{ij}^0 = \begin{cases} 1 & \text{if } j = p, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Since we have no prior knowledge about the label information of the unlabeled data, the reconstruction error associated with each class can be obtained according the classical sparse representation-based classification (SRC) algorithm; so for each unlabeled data, we use the reverse of its reconstruction error associated with a specific class divided by the sum of all the reverse of reconstruction errors associated with each class to represent its initial fuzzy membership associated with that class. We initialize the membership for each unlabeled sample x_k as

$$m_{kj}^0 = \frac{\frac{1}{err_{kj}^0}}{\sum_{i=1}^c \frac{1}{err_{ki}^0}} \quad (\text{for } \forall k \in \{n_l + 1, \dots, n\} \quad \text{and} \quad \forall j \in C) \tag{5}$$

where err_{kj}^0 denotes the reconstruction error of x_k associated with class j , and is obtained according classical SRC algorithm.

m_{ij}^0 and m_{kj}^0 depict the degree of confidence with which a labeled and an unlabeled datum are associated with a specific concept. We construct a membership matrix \mathbf{M}^0 for each datum using the above-defined m_{ij}^0 and m_{kj}^0 with each row representing the memberships of a datum associated with different classes

$$\mathbf{M}^0 = \begin{bmatrix} \mathbf{M}_1^0 \\ \mathbf{M}_u^0 \end{bmatrix} \tag{6}$$

3.2 Membership updating

For convenience, we also use $\mathbf{X}_u \in R^{q \times n_u}$ to denote the data matrix of the unlabeled data, and $\mathbf{X} = [\mathbf{X}_l \mathbf{X}_u] \in R^{q \times n}$ to denote the matrix of all data. Let $\mathbf{X}_{lj} \in R^{q \times n_{lj}}$ be the matrix of the labeled data belonging to class j (n_{lj} is the number of data in class j), we use $\mathbf{X}_j = [\mathbf{X}_{lj} \mathbf{X}_u] \in R^{q \times (n_{lj} + n_u)}$ to denote the data matrix associated with the labeled data in class j and all the unlabeled data, with each column of \mathbf{X}_j representing one datum. We adopt the following l_1 -norm minimization problem to learn the sparse coefficients associated with each unlabeled datum $x_k \in \mathbf{X}_u$ which can be linearly reconstructed by the labeled data and all the unlabeled data except x_k .

$$\min \|x_k - \mathbf{X}\alpha\|^2 + \gamma \|\alpha\|_1 \quad (k \in 1, 2, \dots, n_u) \tag{7}$$

where $\alpha = [\alpha_1, \dots, \alpha_{k-1}, 0, \alpha_{k+1}, \dots, \alpha_n]^T$ is a n -dimensional vector whose k th entry is set to zero (implying that x_k is removed from \mathbf{X}_u), and γ is a parameter to trade-off the sparsity and the reconstruction error.

After α has been obtained, we compute the reconstruction error associated with each class as

$$err_{kj}^t = \left\| x_k - \mathbf{X}\mathbf{W}_j^{t-1}\alpha \right\|_2 \quad (\forall j \in C, \forall k \in \{1, 2, \dots, n_u\} \quad \text{and} \quad \forall t \in \{1, 2, \dots\}) \tag{8}$$

where \mathbf{W}_j^{t-1} is a diagonal matrix associated with class j at the t th iteration, with each diagonal element representing the j th class membership of each datum of \mathbf{X} , and t is an iteration number. err_{kj}^t is the reconstruction error of x_k associated with class j at the t th iteration, and we call it the fuzzy reconstruction error of x_k associated with class j , and $\mathbf{X}\mathbf{W}_j^{t-1}\alpha$ the fuzzy sparse data mean of x_k associated with class j .

The computation of the above class-associated reconstruction errors is similar to classical SRC when all data are labeled, since for computing the reconstruction error associated with class j , the diagonal elements of \mathbf{W}_j^0 are all 1 s for the data belonging to class j , and are 0 s for the data without belonging to class j ; thus, the above formula is reduced to compute class j associated reconstruction error of SRC. This inspires us to introduce the class memberships of the unlabeled data into the reconstruction error computation. The diagonal elements of \mathbf{W}_j^0 are obtained by

$$w_j^0(ii) = m_{ij}^0 \quad (i \in \{1, 2, \dots, n\}, \quad j \in C) \tag{9}$$

We use the following objective function to update the membership of x_u associated with class j

$$\begin{aligned} \min & \sum_{k=1}^{n_u} \sum_{j=1}^c (m_{kj}^t)^r \text{err}_{kj}^t \\ \text{s.t.} & \sum_{j=1}^c m_{kj}^t = 1 \quad (r \in [1, \infty), j \in C, t \in \{1, 2, \dots\}, m_{kj}^t \in [0, 1]) \end{aligned} \tag{10}$$

where r is a fuzzifier which determines the level of class fuzziness, and t is an iteration number. As reported in fuzzy C-means [24] which introduces fuzzy memberships in K-means algorithms, large r results in small memberships and much fuzzy classes, and in the limit $r = 1$, the memberships converge to 0 or 1, which means a crisp partitioning. Equation (10) is by its corresponding Lagrangian

$$J_r = \sum_{k=1}^{n_u} \sum_{j=1}^c (m_{kj}^t)^r \text{err}_{kj}^t + \sum_{k=1}^{n_u} \lambda_k \left(\sum_{j=1}^c m_{kj}^t - 1 \right) \tag{11}$$

By setting the derivative of J_r with respect to m_{kj}^t to 0, we have

$$m_{kj}^t = \frac{1}{\sum_{p=1}^c \frac{\text{err}_{kj}^t - \frac{1}{r-1}}{\text{err}_{kp}^t}} \quad (\forall j \in C \text{ and } \forall k \in \{1, 2, \dots, n_u\}) \tag{12}$$

After the fuzzy memberships have been obtained, we update the fuzzy membership matrix $\mathbf{M}^t = \begin{bmatrix} \mathbf{M}_l^t \\ \mathbf{M}_u^t \end{bmatrix}$, where m_{kj}^t is the k th row and j th column element of \mathbf{W}_u^t .

Then, the diagonal \mathbf{W}_j^t can also be updated as

$$w_j^t(kk) = m_{kj}^t (k \in \{1, 2, \dots, n\}, j \in C, t \in \{1, 2, \dots\}) \tag{13}$$

We continue to update all the class-associated fuzzy reconstruction errors for each datum through Eq. (8), and start a new round membership updating. This process stops if a termination criterion is satisfied. This approach is fuzzy sparse reconstruction error-steered semi-supervised learning (SRE-SSL).

3.3 Membership learning under locality constraints

SR has made great successes in many applications in pattern recognition field, but it employs an overcomplete dictionary, thus neighbors chosen by SR for each datum may be unlikely in its neighborhood. Additionally, SR ignores the local information in the coding process.

Considering that locality is also as important as sparsity, we introduce the locality-constrained linear coding (LLC) [25] in our work by ignoring the shift invariance constraint, and reformulate Eq. (7) as:

$$\min \| \mathbf{x}_k - \mathbf{X}\boldsymbol{\alpha} \|^2 + \gamma \| \mathbf{d} \circ \boldsymbol{\alpha} \|_1 \quad (k \in \{1, 2, \dots, n_u\}) \tag{14}$$

where \circ denotes element-wise multiplication, and \mathbf{d} denotes the weight vector whose j th element is defined as

$$d_{kj} = e^{-\frac{l_{kj}}{\sigma}} \tag{15}$$

where l_{kj} is the Euclidean distance between \mathbf{x}_k and \mathbf{x}_j , and σ is a parameter that is used to adjust the influence of l_{kj} .

Equation (14) is also called a weighted SR which preserves the similarity between the reconstructed datum and its neighbors while seeking the sparse linear representation.

After the sparse coefficients have been obtained, we perform similarly as described in Sect. 3.2 to update the class-associated memberships for each datum. This approach is fuzzy sparse reconstruction error-steered semi-supervised learning with weight (SRE-SSLW).

3.4 Semi-supervised classification

After the final memberships have been obtained, each unlabeled datum \mathbf{x}_k can be classified based on the membership matrix \mathbf{M}^t

$$\text{label}(\mathbf{x}_k) = \arg \max_j m_{kj}^t \quad (\forall j \in C) \tag{16}$$

We summarize the description of the learning process as:

-
- (1) Initialize the membership matrix \mathbf{M}^0 for each unlabeled datum, and set 0 to the iteration step t ;
 - (2) Do the following until termination:
 $\{ t = t + 1$; compute the fuzzy reconstruction errors associated with each class for each unlabeled datum based on \mathbf{M}^{t-1} and Eq. (8); update the membership m_{kj}^t for \mathbf{M}^t according to Eq. (12);
 - (3) Classify the unlabeled data according to Eq. (16).
-

The stopping condition can be defined differently; in this paper, we stop to update the fuzzy memberships when a given number of iterations are met.

The algorithm classifies each datum based on the assumption that data are bound to each class by means of a membership function, which represents the fuzzy behavior of this algorithm. To do that, we construct a membership matrix \mathbf{M} whose entries lie in $[0, 1]$, and use the entries to represent the degrees that a datum belongs to each class.

When computing the reconstruction error associated with each class, the classical SRC ignores the labeled data belonging to the other classes. Unlike this method, we incorporate the label information. Since Eq. (8) can be reduced to SRC, the computation of the construction error can be considered as a specific case of Eq. (8). Furthermore, for a specific class j , if we take the labeled data into account and assign zeros to the labeled data from classes that are different from class j , these labeled data make no contribution to the reconstruction error associated with class j .

4 Experimental comparison

We do experiments on commonly used benchmark face image datasets to evaluate the performance of the proposed approach, and make comparisons with the state-of-the-art approaches.

4.1 Datasets

The image datasets include extended YaleB¹, ORL², the Yale³, FERET⁴, and AR⁵. The dataset descriptions are given as follows, and sample images are shown in Fig. 1.

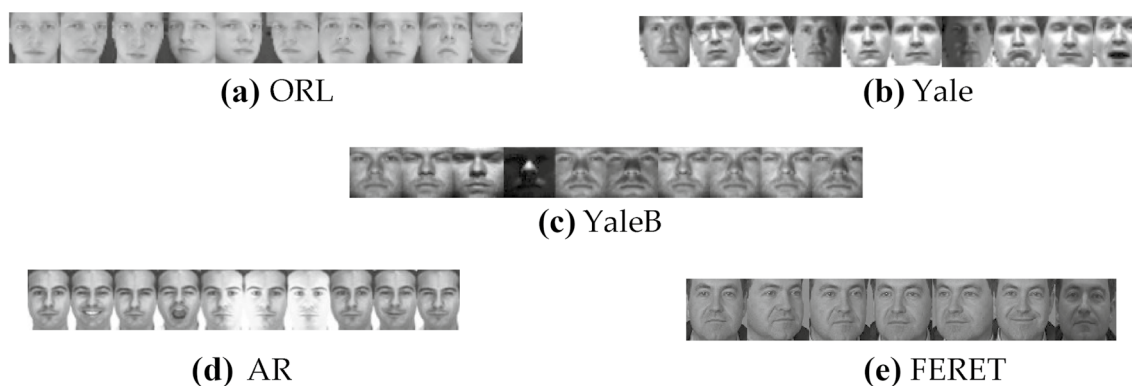


Fig. 1 Sample images from each image set

1. The extended YaleB contains images of 38 individuals with about 64 images for each individual. All images are cropped and resized to 32×32 pixels.
2. The ORL contains ten different face images for each of 40 subjects. These images were collected at different times and under different conditions, such as different lighting, facial expressions, and facial details. All images are cropped and resized to 32×32 pixels.
3. The Yale face dataset contains 165 grayscale images of 15 individuals, and each individual has 11 images under different facial expressions or configurations. Each image is cropped and resized to 32×32 pixels.
4. AR contains 4000 color frontal images taken from 126 subjects at two separate sessions with different occlusions illumination variation and facial expression, and is used for gender recognition. We use a subset consisting of 50 male and 50 female subjects with total number of 2600 images. Each class consists of 26 images with different facial expressions and illumination conditions, and each image is resized to 60×43 pixels.
5. The FERET face image database has become a standard database for testing and evaluating state-of-the-art face recognition algorithms. The proposed method is tested on a subset of the FERET database. This subset includes 700 images of 100 individuals. (Each one has seven images.) This subset involves variations in facial expression, illumination, and pose. In our experiment, the facial portion of each original image is automatically cropped based on the location of eyes, and the cropped image is resized to 40×40 pixels and further reprocessed by histogram equalization.

4.2 Experimental setup

In order to evaluate the performance in detail on each dataset, we randomly choose 10 to 60 percent of the total images from each class as labeled ones, and consider the remaining images as unlabeled ones by ignoring their label

information. Each experiment is repeated ten times for each algorithm, and the mean accuracies of each algorithm are reported.

Extensive experiments on face images have been done by the recent study MLRR [15] for performance comparison with the other state-of-the-art SSL models. These models include the KNN approach utilizing Gaussian kernel to calculate the graph edge weights, the local spline regression-based graph [26] and label propagation through linear neighborhoods [5] taking into account the local and global characteristics in construction of the Laplacian matrix, the sparse representation technique-based graph edge weight calculation, and the low-rank representation and its several variants. The mean accuracies demonstrate that MLRR outperforms the other models almost all the time; so we choose MLRR and the state-of-the-art sparse representation related models for performance comparison, and do not repeat the experiments for the models compared with MLRR.

We compare our approach with sparsity-induced similarity measure for label propagation (SIS) [4], label propagation through linear neighborhoods (LNP) [5], label propagation through sparse neighborhood (LPSN) [13] and manifold low-rank representation (MLRR) [15] and fuzzy K-SVD (F-K-SVD) [21]. The first four methods are based on semi-supervised learning and the last method achieves the fuzzy sparse signal representations according to dictionary learning. We follow the parameter setup in [15] for MLRR with N_i , γ , α , and β being set to 20, 1, 0.01, and 0.01, respectively, and follow the parameter setup for the compared models SIS [4], LPSN [13], LNP [5], and F-K-SVD [21] for their optimal parameter choice.

We report the results of our algorithm on the ORL and Yale datasets under four different conditions: **SRE-SSL1**, **SRE-SSL2**, **SRE-SSLW1**, and **SRE-SSLW2**. **SRE-SSL1**: the sparse coefficients and the fuzzy memberships are obtained through Eqs. (7) and (12) separately, and the fuzzy memberships are introduced to the fuzzy

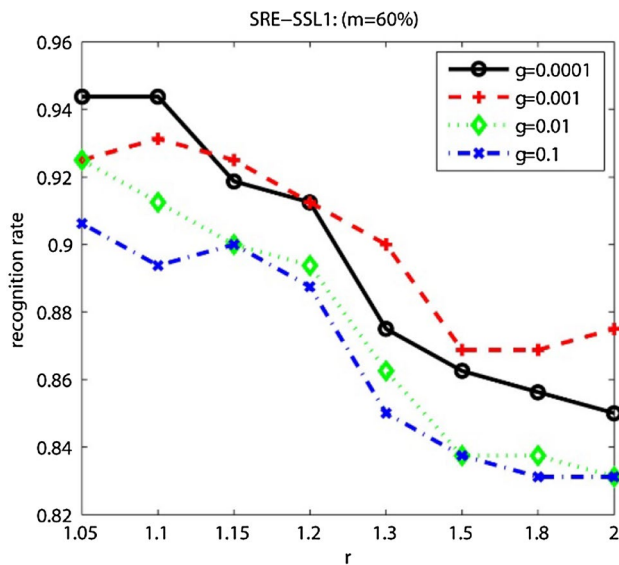


Fig. 2 Recognition rate on ORL under different parameter values

reconstruction error calculation in each iteration; **SRE-SSL2**: For each unlabeled datum, after the fuzzy reconstruction error associated with each class are obtained, we assign 1 to the fuzzy membership of a specific class with which the reconstruction error is associated, and 0 to the fuzzy membership of the other classes. Moreover, the weights are introduced to the sparse coefficients when learning the dictionary for each unlabeled datum in Eq. (14), and then two conditions form and are called **SRE-SSLW1** and **SRE-SSLW2**, respectively.

4.3 Experimental results

The parameters should be predetermined before experimental comparison. In order to obtain the optimal parameter values, we do experiments when the proportion of the labeled data is fixed to 60% on each dataset. Three parameters should be determined for **SRE-SSLW1**, two are needed for **SRE-SSL1** and **SRE-SSLW2**, and one is needed for **SRE-SSL2**. Based on the previous works on dictionary learning and fuzzy C-means algorithm, we assign a value set 0.0001, 0.001, 0.01, 0.1 to γ and a value set 2, 1.8, 1.5, 1.3, 1.2, 1.1, 1.05 to r . We choose ORL image dataset, and conduct **SRE-SSL1** under different parameter value combinations in the case that 60% data in each class are randomly selected and treated as labeled data. The experiments are repeated ten times, and mean accuracy rates are obtained. Figure 2 demonstrates the performance of **SRE-SSL1** on ORL under different parameter setup when 60% data in each class are randomly selected and treated as labeled data. It is obvious that the classification accuracy is sensitive to the value of γ

and r , and the optimal values can be obtained when γ is set to 0.0001 and r is set to 1.1.

We implement **SRE-SSL1** on the Yale image set under the optimal parameter setup, and show the convergence properties of the fuzzy memberships. We stop membership updating after ten training iterations. After termination, we randomly choose one unlabeled sample, and show its memberships associated with each class in Fig. 3. We use one subfigure to demonstrate its convergence property of membership associated with each class. The memberships show different convergence properties. Some converge correctly at the very beginning iteration, and some converge correctly after several iterations. There are still some memberships that do not converge before terminating the iteration. The memberships of all correctly classified samples converge, and the memberships of wrongly classified samples do not converge.

Since there are three parameters that should be determined for **SRE-SSLW1**, we assign a value set 0.01, 0.1, 1, 10, 100 to σ , and do experiments under different value combination conditions. Tenfold cross-validation is implemented, and based on the mean recognition rates, the optimal value combination of γ , r , and σ is obtained as (0.1, 1.05, 1) for ORL, Yale, YaleB, and AR, and the optimal parameter combination is (0.1, 1.2, 1) for FERET.

We conduct **SRE-SSLW1** on the Yale image set under the optimal parameter setup, and show the membership convergence properties in Fig. 4. It is obvious that the convergence properties of the memberships obtained by **SRE-SSLW1** are similar to **SRE-SSL1**. We do experiments and obtain the optimal parameter value 0.0001 for γ in **SRE-SSL2**, and the optimal parameter value combination (0.1, 1) for γ, σ in **SRE-SSLW2**. Extensive experiments under different percentages of labeled data have been done on each of the five image sets, and the mean accuracies together with the corresponding standard deviations over ten trials are shown in Tables 1 and 2. All the results are obtained under the optimal parameter setup for each algorithm, and we can easily observe from the results that **SRE-SSL1** outperforms **SRE-SSL2**, and **SRE-SSLW1** outperforms **SRE-SSLW2** under all conditions. This demonstrates that data are bound to each class by means of a membership function which represents the fuzzy behavior of both **SRE-SSL1** and **SRE-SSLW1**. Similar to K-means algorithm, **SRE-SSL2** and **SRE-SSLW2** assign each datum to a specific class in each iteration based on its membership associated with each class. This manipulation causes error and enlarges the error in the following iterations, and does harm to the performance of **SRE-SSL2** and **SRE-SSLW2**.

Face images are manifold data, thus considering the local structure of data is beneficial to image recognition.

Fig. 3 Membership convergence property of SRE-SSL1

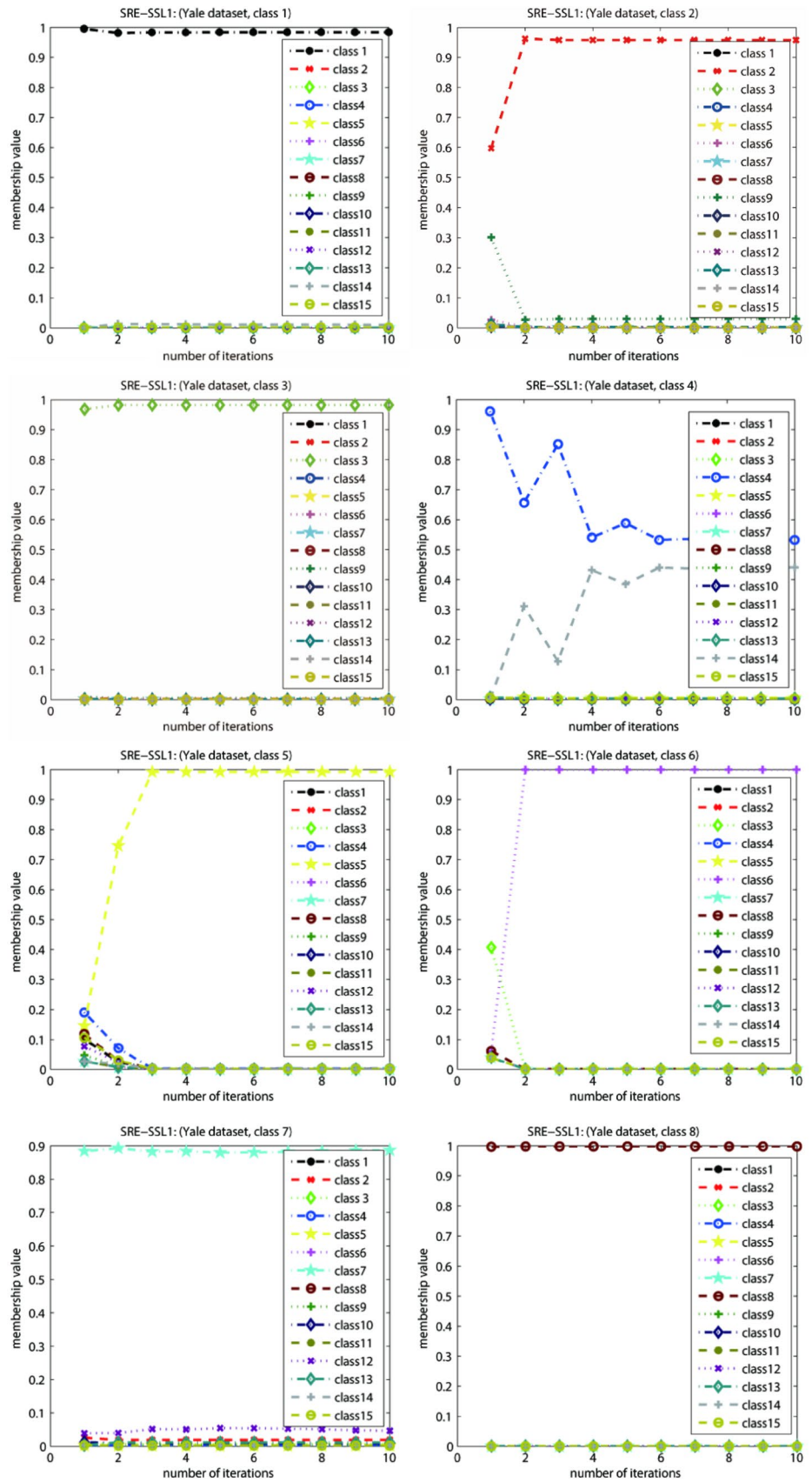


Fig. 3 (continued)

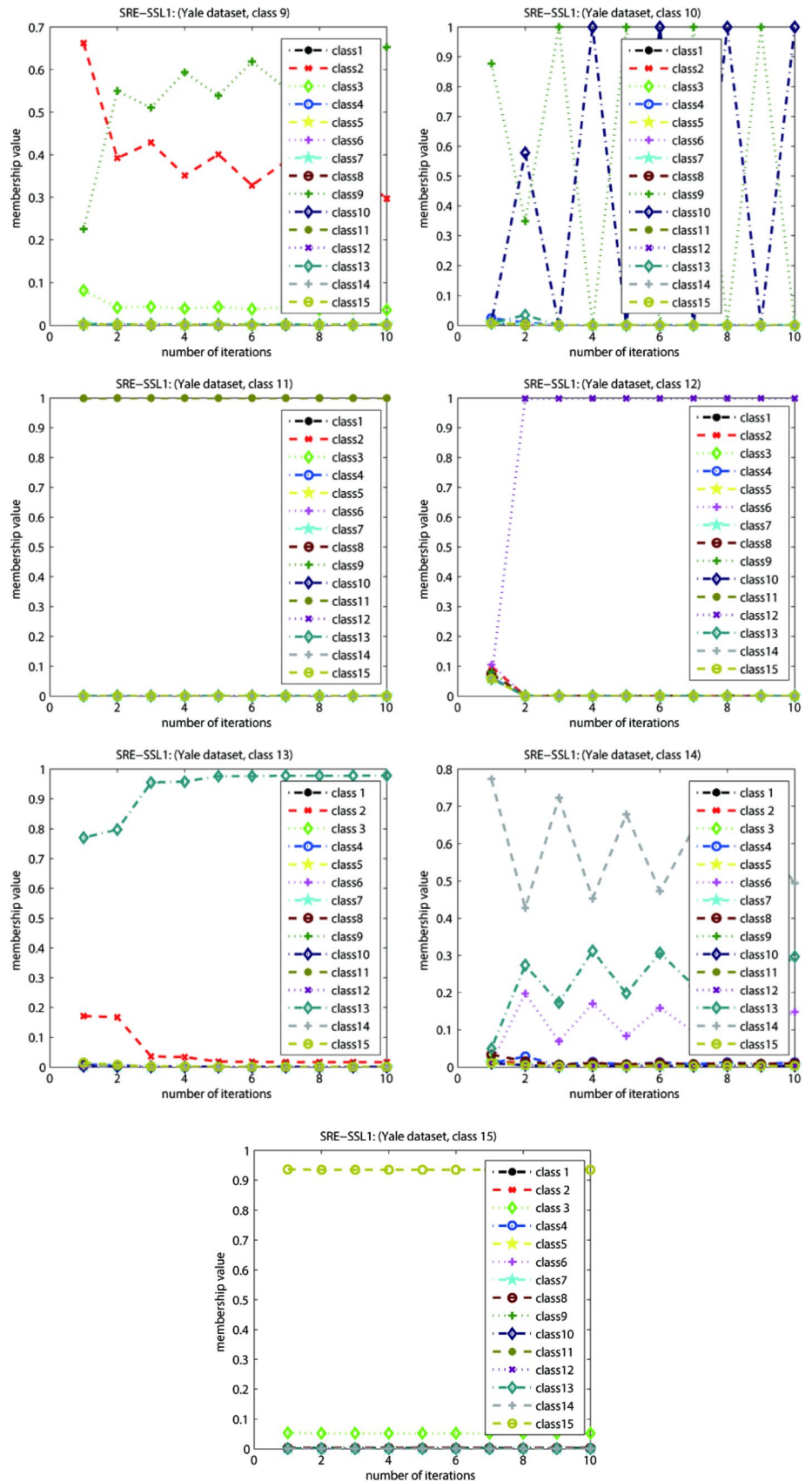


Fig. 4 Membership convergence property of SRE-SSLW1

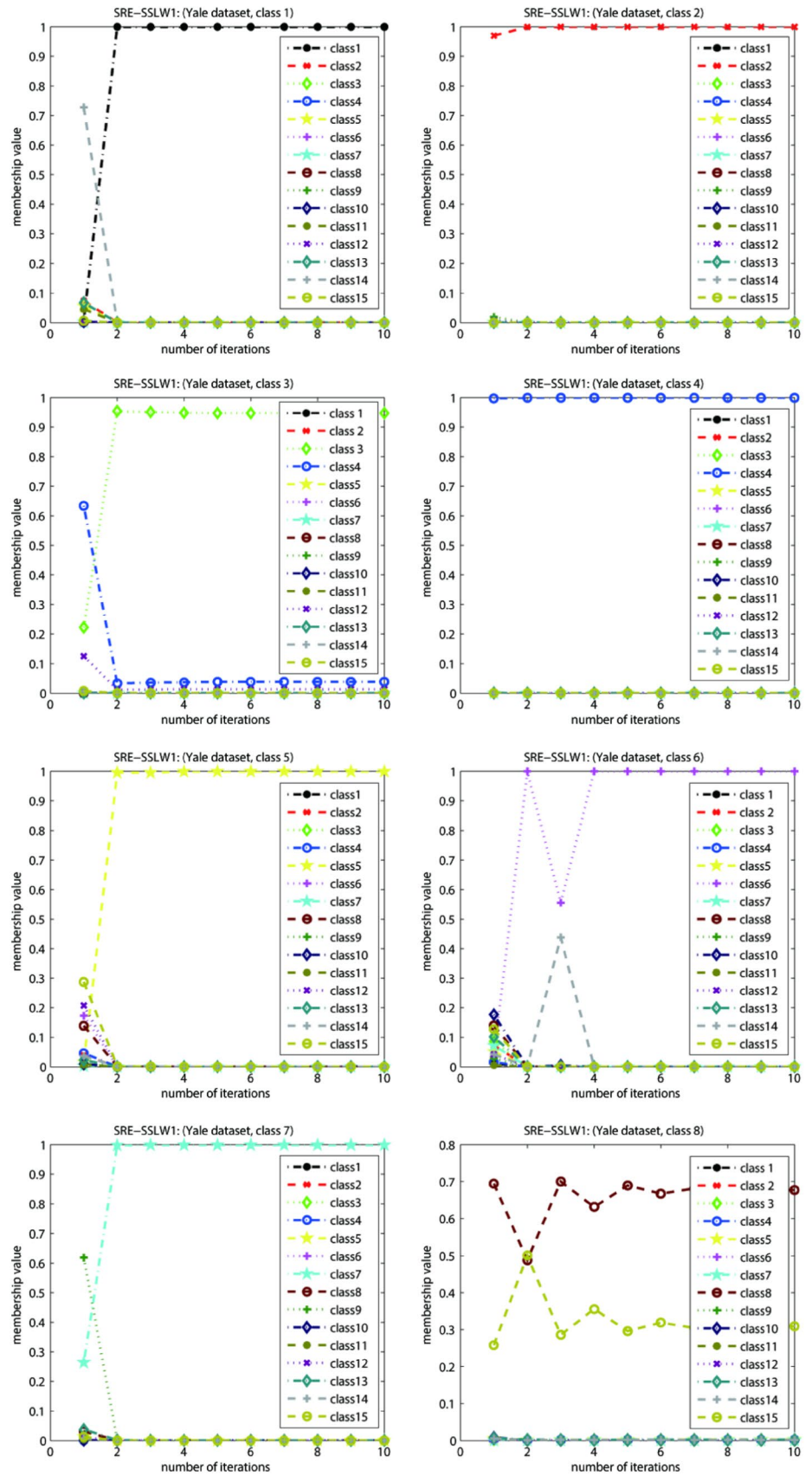


Fig. 4 (continued)

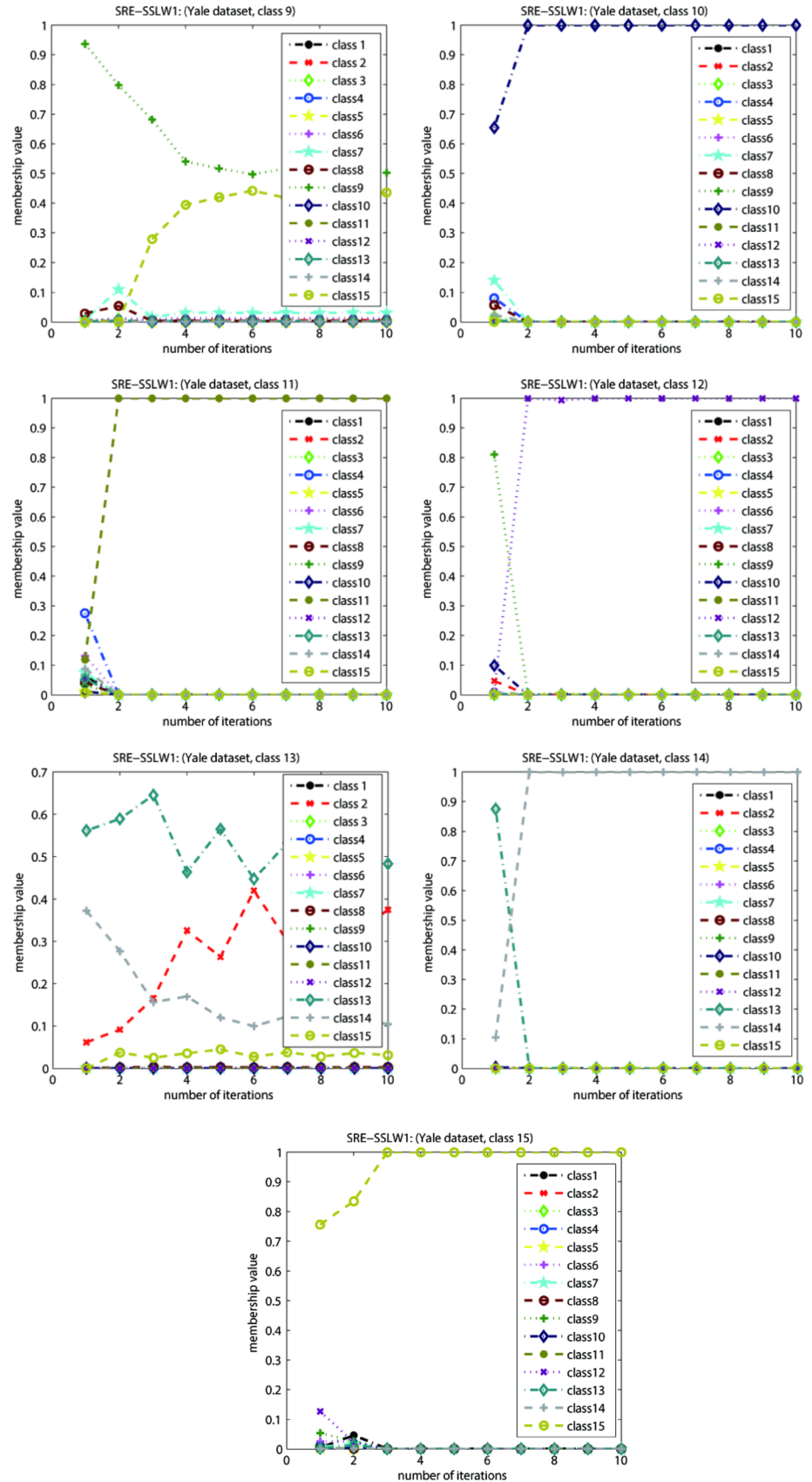


Table 1 Results on ORL and Yale datasets (mean \pm SD%)

	60%	50%	40%	30%	20%	10%
ORL						
LNP	89.50 \pm 1.61	85.65 \pm 2.91	82.29 \pm 1.42	74.96 \pm 4.33	68.50 \pm 2.37	58.25 \pm 1.78
SIS	94.38 \pm 2.38	92.40 \pm 1.68	89.13 \pm 1.77	82.86 \pm 3.69	74.16 \pm 1.98	54.69 \pm 2.03
LPSN	89.81 \pm 2.96	88.55 \pm 2.10	87.92 \pm 1.98	84.75 \pm 3.26	80.19 \pm 1.70	70.94 \pm 2.90
MLRR	96.90 \pm 1.59	94.60 \pm 1.22	94.33 \pm 1.56	90.50 \pm 2.53	85.78 \pm 2.44	73.14 \pm 2.36
SRE-SSL1	96.13 \pm 1.58	95.05 \pm 1.04	93.88 \pm 2.44	89.86 \pm 3.74	83.19 \pm 2.38	61.53 \pm 2.25
SRE-SSL2	90.13 \pm 2.30	85.30 \pm 2.68	79.25 \pm 2.11	68.93 \pm 3.85	52.56 \pm 3.77	23.83 \pm 2.71
SRE-SSLW1	96.31 \pm 1.60	96.10 \pm 1.37	94.96 \pm 1.97	91.96 \pm 2.34	86.25 \pm 2.61	69.75 \pm 3.65
SRE-SSLW2	91.06 \pm 2.17	87.15 \pm 3.94	78.88 \pm 2.59	68.89 \pm 3.21	55.13 \pm 1.58	37.53 \pm 3.58
Yale						
LNP	61.33 \pm 3.67	57.60 \pm 5.36	51.71 \pm 2.98	46.50 \pm 4.49	43.26 \pm 3.19	35.67 \pm 3.47
SIS	66.17 \pm 7.50	60.40 \pm 7.65	49.14 \pm 12.35	44.17 \pm 14.59	32.89 \pm 13.72	19.67 \pm 10.84
LPSN	62.00 \pm 8.45	62.40 \pm 5.58	57.81 \pm 4.28	53.58 \pm 4.14	44.59 \pm 3.18	39.40 \pm 3.96
MLRR	49.33 \pm 6.30	48.53 \pm 5.56	43.14 \pm 5.12	40.42 \pm 4.59	37.48 \pm 4.08	34.27 \pm 4.68
SRE-SSL1	80.67 \pm 4.10	80.00 \pm 3.82	65.43 \pm 5.48	53.08 \pm 5.11	37.85 \pm 5.46	28.20 \pm 4.62
SRE-SSL2	71.33 \pm 6.42	69.07 \pm 4.86	51.62 \pm 6.73	44.67 \pm 6.41	34.15 \pm 4.44	21.00 \pm 2.67
SRE-SSLW1	82.50 \pm 4.73	81.60 \pm 4.11	71.14 \pm 4.49	64.58 \pm 7.26	48.67 \pm 5.06	39.80 \pm 6.73
SRE-SSLW2	78.33 \pm 4.58	72.80 \pm 4.63	60.48 \pm 5.12	50.50 \pm 5.87	40.07 \pm 4.83	27.47 \pm 4.60

Table 2 Results on AR, YaleB, and FERET datasets (mean \pm SD%)

	60%	50%	40%	30%	20%	10%
AR						
LNP	73.13 \pm 1.91	69.66 \pm 1.97	64.58 \pm 1.03	54.32 \pm 0.15	48.42 \pm 1.30	29.94 \pm 1.85
SIS	91.53 \pm 2.39	92.26 \pm 1.40	89.03 \pm 2.09	87.76 \pm 0.85	78.60 \pm 10.97	39.85 \pm 19.95
LPSN	88.83 \pm 0.68	88.57 \pm 1.91	86.50 \pm 1.29	81.94 \pm 1.49	79.56 \pm 1.45	62.65 \pm 3.12
MLRR	97.87 \pm 0.58	97.63 \pm 0.46	96.83 \pm 0.63	95.82 \pm 0.40	93.33 \pm 0.55	75.06 \pm 2.02
SRE-SSLW1	98.33 \pm 0.35	98.23 \pm 0.90	97.78 \pm 0.80	97.18 \pm 0.18	97.02 \pm 0.79	74.11 \pm 4.60
YaleB						
LNP	81.16 \pm 0.89	80.22 \pm 0.56	76.21 \pm 1.29	72.86 \pm 0.48	67.87 \pm 0.88	57.72 \pm 0.97
SIS	82.79 \pm 2.13	83.61 \pm 2.76	76.09 \pm 0.93	73.49 \pm 4.98	70.37 \pm 5.39	42.11 \pm 18.09
LPSN	91.07 \pm 0.87	89.78 \pm 0.83	88.38 \pm 1.07	86.47 \pm 0.65	82.98 \pm 1.91	74.72 \pm 2.03
MLRR	98.34 \pm 0.16	98.12 \pm 0.24	97.30 \pm 0.39	95.47 \pm 0.59	92.69 \pm 0.84	86.07 \pm 1.01
SRE-SSLW1	99.06 \pm 0.10	98.93 \pm 0.09	98.96 \pm 0.14	98.52 \pm 0.23	96.05 \pm 0.74	93.35 \pm 1.87
FERET						
LNP	46.53 \pm 0.38	45.53 \pm 1.18	38.33 \pm 0.89	32.12 \pm 0.91	21.85 \pm 0.97	22.02 \pm 0.93
SIS	54.00 \pm 2.11	50.77 \pm 1.24	45.50 \pm 3.23	35.34 \pm 3.07	22.27 \pm 3.18	19.70 \pm 3.40
LPSN	53.17 \pm 0.72	50.83 \pm 0.57	44.18 \pm 2.24	36.72 \pm 1.00	24.30 \pm 1.27	24.20 \pm 1.17
MLRR	60.53 \pm 0.81	60.03 \pm 1.86	54.05 \pm 1.18	45.44 \pm 0.67	29.90 \pm 1.90	29.90 \pm 1.24
SRE-SSLW1	75.13 \pm 0.84	72.70 \pm 0.94	66.53 \pm 1.39	56.74 \pm 1.78	43.93 \pm 1.19	37.98 \pm 1.29

Since both sparsity and locality are important for classification, **SRE-SSLW1** introduces a weight matrix to the sparse reconstruction of each datum, which characterizes the local structure of data. The results indicate that **SRE-SSLW1** outperforms the other algorithms almost all the time except that it only loses two times under 10% of the training data. **SRE-SSLW1** wins MLRR three times under 10% training data condition. In order to make a fair performance

comparison, we obtain the average accuracies of LNP, SIS, LPSN, MLRR, F-K-SVD, and **SRE-SSLW1** on the five datasets with values 40.72, 35.20, 54.38, 59.69, 42.80, and 63.00, respectively. The average values also indicate that **SRE-SSLW1** performs the best.

We also obtain the ranks of the six approaches on each image set, and compare their average ranks. According to the averaged accuracies of ten trails on each face database, the

Table 3 Ranks and averaged ranks of each algorithm under 10% ratio

	ORL	Yale	AR	YaleB	FERET	Average rank
LNP	5	3	6	4	4	4.4
SIS	6	6	4	5	5	5.2
LPSN	3	2	3	3	3	2.8
MLRR	1	5	1	2	2	2.2
WSRE-SSL1	2	1	2	1	1	1.4

best one gets a rank of 1, and the second best one gets rank of 2, and so on. If approaches perform equally well, average ranks are assigned to them. Table 3 reports both the ranks and the average ranks of each algorithm, and **SRE-SSLW1** also performs the best according to the average rank.

5 Conclusion

This paper introduces a semi-supervised learning approach to face recognition. The sparse representation techniques are utilized to capture the sparse relationship among data, and each unlabeled datum is linearly and sparsely reconstructed by a small number of labeled data together with a large number of unlabeled ones. When calculating the class-associated reconstruction errors for each unlabeled datum, the determined or fuzzy label information of data is considered. Since labeled data are bounded to specific classes, and unlabeled ones are not clearly bounded to any specific classes, we utilize membership functions to characterize the unspecific class properties of the unlabeled ones. The fuzzy membership is a function of the reconstruction errors and is introduced to the reconstruction error calculation. It is repeatedly updated after new reconstruction errors are obtained until a termination criterion is satisfied. Each unlabeled datum is classified based on its memberships associated with each class. Considering that face images are manifold data and both sparsity and locality are essential to image recognition, we utilize a weighted sparse representation technique to obtain the sparse coefficients so as to characterize the embedded data locality. The sparse coefficients together with the fuzzy memberships play important roles in the reconstruction error calculation, and experimental results indicate that locality is also helpful to face recognition. In the future, we will try to consider utilizing the time-saving collaborative representation techniques to obtain the reconstruction coefficients.

Acknowledgements This work was supported by the National Natural Science Foundation of China (61702310 and 61772322).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Human and animal rights This article does not contain any studies involving human participants performed by any of the authors.

References

- Zhou, D.Y., Bousquet, O., Lal, T.N., Weston, J., Olkoph, B.S.: Learning with local and global consistency. *Adv. Neural Inf. Process. Syst.* **16**(3), 321–328 (2004)
- Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**(5500), 2323–2326 (2000)
- Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Supervised dictionary learning. In: *Computer Science*, pp. 1–8 (2008)
- Cheng, H., Liu, Z., Yang, J.: Sparsity induced similarity measure for label propagation. In: *Proceedings of the IEEE 12th international conference on computer vision*, pp. 317–324 (2009)
- Wang, F., Zhang, C.S.: Label propagation through linear Neighborhoods. *IEEE Trans. Knowl. Data Eng.* **20**(1), 55–67 (2008)
- Cheng, B., Yang, J., Yan, S.C., Fun, Y., Huang, T.S.: Learning with l_1 -graph for image analysis. *IEEE Trans. Image Process.* **19**(4), 858–866 (2010)
- Muller, K., Mika, S., Riitsch, G., Tsuda, K., Scholkopf, B.: An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Netw.* **12**, 181–201 (2001)
- Tang, J., Hua, X.-S., Song, Y., Qi, G.-J., Wu, X.: Kernel-based linear neighborhood propagation for semantic video annotation. In: *Pacific-Asia Conference on Advances in Knowledge Discovery and Data*, Nanjing, China, pp. 793–800 (2007)
- Kobayashi, T.: Kernel-based transition probability toward similarity measure for semi-supervised learning. *Pattern Recognit.* **47**, 1994–2010 (2014)
- Rakotomamonjy, A., Bach, F.R., Canu, S., Grandvalet, Y.: SimpleMKL. *J. Mach. Learn. Res.* **9**, 2491–2521 (2008)
- Wang, S., Huang, Q., Jiang, S., Tian, Q.: (SMKL)-M-3: scalable semi-supervised multiple kernel learning for real-world image applications. *IEEE Trans. Multimedia* **14**(4), 1259–1274 (2012)
- Zhao, Z.Q., Glotin, H., Xie, Z., Gao, J., Wu, X.: Cooperative sparse representation in two opposite directions for semi-supervised image annotation. *IEEE Trans. Image Process.* **21**(9), 4218–4231 (2012)

13. Zang, F., Zhang, J.-S.: Label propagation through sparse neighborhood and its applications. *Neurocomputing* **97**, 267–277 (2012)
14. Zhuang, L., Gao, H., Lin, Z., Ma, Y., Zhang, X., Yu, N.: Nonnegative low rank and sparse graph for semi-supervised learning. In: *Proceedings of 25th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2012)
15. Peng, Y., Lu, B.-L., Wang, S.: Enhanced low-rank representation via sparse manifold adaption for semi-supervised learning. *Neural Netw.* **65**, 1–17 (2015)
16. Fan, M., Gu, N., Qiao, H., Zhang, B.: Sparse regularization for semi-supervised classification. *Pattern Recognit.* **44**, 1777–1784 (2011)
17. Shang, F., Jiao, L.C., Liu, Y., Tong, H.: Semi-supervised learning with nuclear norm regularization. *Pattern Recognit.* **6**, 2323–2336 (2013)
18. Zhang, Z., Zhao, M., Chow, T.W.S.: Graph based constrained semi-supervised learning framework via label propagation over adaptive neighborhood. *IEEE Trans. Knowl. Data Eng.* **27**(9), 2362–2376 (2015)
19. Wang, Z.H., Yoon, S., Xie, S.J., Lu, Y., Park, D.S.: Visual tracking with semi-supervised online weighted multiple instance learning. *Vis. Comput.* **32**, 307–320 (2016)
20. Pei, X., Lyu, Z., Chen, C., Chen, C.: Manifold adaptive label propagation for face clustering. *IEEE Trans. Cybernet.* **45**(8), 1681–1691 (2015)
21. Song, X., Liu, Z.: A fuzzy adaptive K-SVD dictionary algorithm for face recognition. *Appl. Mech. Mater.* **347**, 3797–3803 (2013)
22. Zhu, X., Ghahramani, Z., Lafferty, J.: Semi-supervised learning using gaussian fields and harmonic functions. In: *Proceedings of the 20th International Conference on Machine Learning*, vol 20, pp 912–919 (2003)
23. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009)
24. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York (1981)
25. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: *Proceedings of 23th IEEE Conference on Computer Vision and Pattern Recognition*, pp 3360–3367 (2010)
26. Xiang, S., Nie, F., Zhang, C.: Semi-supervised classification via local spline regression. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 2039–2053 (2010)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Li Liu received the Ph.D. degree from Shandong University, where she was involved in computer vision research. She is currently an Associate Professor of information science and engineering with Shandong Normal University. Her research interests include image processing, computer vision, and pattern recognition.



Siqi Chen is a masters student in the School of Information Science and Engineering in Shandong Normal University. She received her bachelors degree in computer science and technology from Shandong Normal University. Her research interests include machine learning and 3D model retrieval.



Xiuxiu Chen is a masters student in the School of Information Science and Engineering in Shandong Normal University. She received her bachelors degree from Shandong Jiaotong University. Her research interests include machine learning and image inpainting.



Tianshi Wang is a masters student in the School of Information Science and Engineering in Shandong Normal University. He received his B.S. degree from Shandong Normal University. His research interests include natural language processing, cross-media retrieval, and computer vision.



Long Zhang is a masters student in the School of Information Science and Engineering in Shandong Normal University. He received his bachelors degree in computer science from Shandong Normal University. His research interests include machine learning and cross-media retrieval.