Structure Preserving Low-Rank Representation for Semi-supervised Face Recognition

Yong Peng[1], Suhang Wang[2], Shen Wang[2], and Bao-Liang Lu[1,3,⋆]

[1] Center for Brain-Like Computing and Machine Intelligence,
Department of Computer Science and Engineering,
Shanghai Jiao Tong Unviersity, Shanghai 200240 China
[2] Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, 48109 USA
[3] MOE-Microsoft Key Lab. for Intelligent Computing and Intelligent Systems
Shanghai Jiao Tong Unviersity, Shanghai 200240 China
bllu@sjtu.edu.cn

Abstract. Constructing an informative and discriminative graph plays an important role in the graph based semi-supervised learning methods. Among these graph construction methods, low-rank representation based graph, which calculates the edge weights of both labeled and unlabeled samples as the low-rank representation (LRR) coefficients, has shown excellent performance in semi-supervised learning. In this paper, we additionally impose twofold constraints (*local affinity* and *distant repulsion*) on the LRR graph. The improved model, termed structure preserving LRR (SPLRR), can preserve the local geometrical structure but without distorting the distant repulsion property. Experiments are taken on three widely used face data sets to investigate the performance of SPLRR and the results show that it is superior to some state-of-the-art semi-supervised graphs.

Keywords: Structure preserving, Low-rank representation, Semi-supervised learning, Face recognition.

1 Introduction

Recently, semi-supervised learning (SSL) has received increasing attention because it can utilize both limited labeled samples and rich yet unlabeled samples. The currently available semi-supervised methods can be roughly categorized into four groups: generative models, low-density separation models, heuristic models and graph-based models. In this paper, we focus our work on graph-based SSL due to its empirical success in practice and computational efficiency.

Graph-based SSL relies on using a graph $G = (V, E, W)$ to represent data structure, where V is a set of vertices in which each vertex represents a data point, $E \subseteq V \times V$ is a set of edges connecting related vertices and W is an adjacency matrix recording the pairwise weights between vertices. Usually, the graph

⋆ Corresponding author.

M. Lee et al. (Eds.): ICONIP 2013, Part II, LNCS 8227, pp. 148–155, 2013.

is constructed using relationship of domain knowledge or similarity of samples. Upon the graph is constructed, each sample spreads its label information to its neighbors until a global stable state is achieved on the whole data set. Thus, both labeled and unlabeled samples remarkably affect the construction of graphs and how to construct a good graph for representing the data structure is critical for graph-based SSL. Recently, some commonly used graphs have been well investigated, such as k nearest neighbors graph [10], graph for label propagation based on linear neighborhoods (LNP) [11], ℓ^1 graph [5], sparse probability graph (SPG) [6] and so on.

The ℓ^1 graph is motivated by which each datum can be reconstructed by the sparse linear superposition of the training data [5] and the sparse reconstruction coefficients are derived by solving an ℓ^1 optimization problem. Differing from the sparse representation which enforces the representation coefficients to be sparse, the semi-supervised low-rank representation graph (LRR) was proposed for pattern classification [12]. However, the low rankness constraint can only capture the global mixture of subspaces structure while ignoring the local structure of data. To compensate the drawback of LRR graph, we propose a structure preserving low-rank representation based graph, which is imposed on twofold constraints: local affinity and distant repulsion. Therefore, the proposed structure preserving low-rank representation can properly preserve the local affinity structure without distorting the distant repulsion property.

The remainder of this paper is organized as follows. We present a brief review of low-rank representation in section 2. In section 3, we propose the formulation of structure preserving low-rank representation (SPLRR) model and its implementation which is based on the inexact ALM algorithm. Section 4 shows the semi-supervised classification method used in this paper. Experiments on three widely used face databases for evaluating the performance of SPLRR are illustrated in section 5. Conclusion is given in section 6.

2 Low-Rank Representation

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ be a set of data points in d-dimensional space. We try to represent each sample in $\mathbf{X}$ based on the dictionary $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_m] \in \mathbb{R}^{d \times m}$ using $\mathbf{X} = \mathbf{AZ}$, where $\mathbf{Z}$ is the representation coefficients matrix. When the dictionary $\mathbf{A}$ is over-complete, there will be many solutions to this problem. LRR seeks a lowest-rank solution by solving the following problem [8] (we use the data matrix $\mathbf{X}$ itself as dictionary):

$$\min_{\mathbf{z}} \quad \text{rank}(\mathbf{Z}), \ \ s.t. \ \mathbf{X} = \mathbf{XZ}. \tag{1}$$

The optimal solution to (1) is called the "lowest-rank representations" of data $\mathbf{X}$ w.r.t. a dictionary $\mathbf{X}$. However, this problem is NP-hard to solve due to the discrete nature of *rank* function. Fortunately, we can convert (1) to following convex optimization problem instead based on the work of [3]:

$$\min_{\mathbf{Z}} \quad ||\mathbf{Z}||_\star, \quad s.t. \ \mathbf{X} = \mathbf{XZ}, \tag{2}$$

where $||\cdot||_\star$ denotes the nuclear norm of a matrix [2], i.e., the sum of the singular values of the matrix.

In real-world applications, data points are often noisy or even grossly corrupted. Therefore, the corrupted data can be separated to two parts, i.e., $\mathbf{X} = \mathbf{XZ} + \mathbf{E}$. Thus, the affinity matrix $\mathbf{Z}$ can be obtained by solving the following problem:

$$\min_{\mathbf{Z},\mathbf{E}} \quad ||\mathbf{Z}||_\star + \lambda||\mathbf{E}||_\ell, \quad s.t.\ \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \tag{3}$$

where $||\cdot||_\ell$ can be the $\ell_{2,1}$ or ℓ_1 norm (in this paper we choose ℓ_1 norm). The optimal solution $\mathbf{Z}^\star$ to problem (3) can be obtained via the Inexact Augmented Lagrange Multiplier Method (ALM) [7].

3 Structure Preserving Low-Rank Representation

In this section, we propose the structure preserving low-rank representation (SPLRR) model as well as its solution based on Inexact Augmented Lagrange Multiplier Method [7]. SPLRR can properly preserve the local affinity structure without distorting the distant repulsion property. The local affinity indicates the local neighborhood correlation, which means that if $\mathbf{x}_i$ and $\mathbf{x}_j$ are close in the original data space, their corresponding representation coefficients $\mathbf{z}_i$ and $\mathbf{z}_j$ should be also close in the transformed space. The distant repulsion property is inspired by the elastic embedding [4], which enforces the corresponding representation coefficients of distant data points in the original space to be kept distant in the transformed space.

3.1 Formulation of SPLRR

We first introduce the two constraints: *local affinity* and *distant repulsion*.

• Local affinity. To preserve the local geometrical structure in the coefficient space, one may naturally hope that, if two data points $\mathbf{x}_i$ and $\mathbf{x}_j$ are close in the intrinsic manifold, their corresponding representation coefficients $\mathbf{z}_i$ and $\mathbf{z}_j$ should also be close to each other. This can be viewed as manifold assumption for smoothness transition. Several methods can achieve this manifold-like property and in this work we choose the graph regularization term which is similar to graph regularized non-negative matrix factorization (GNMF) in [1]:

$$\begin{aligned}\min_{\mathbf{Z}} \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}||\mathbf{z}_i - \mathbf{z}_j||^2 &= \sum_{i=1}^{n} d_{ii}\mathbf{z}_i^T\mathbf{z}_i - \sum_{i,j=1}^{n} w_{ij}\mathbf{z}_i^T\mathbf{z}_j \\ &= \mathrm{Tr}(\mathbf{ZDZ}^T) - \mathrm{Tr}(\mathbf{ZWZ}^T) = \mathrm{Tr}(\mathbf{ZL}_1\mathbf{Z}^T)\end{aligned} \tag{4}$$

where $\mathbf{D}$ is a diagonal matrix whose entries are column (or row, since $\mathbf{W}$ is symmetric) sums of $\mathbf{W}$, $d_{ii} = \sum_j w_{ij}$. We can compute the graph Laplacian $\mathbf{L}_1 = \mathbf{D} - \mathbf{W}$, w_{ij} is the measure of affinity between $\mathbf{x}_i$ and $\mathbf{x}_j$ in the original space. Here, we use the 'HeatKernel' formulation:

$$w_{ij} = \begin{cases} \exp(-||\mathbf{x}_i - \mathbf{x}_j||^2/2t^2), & \text{if } \mathbf{x}_i \in \mathcal{N}_{k1}(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \mathcal{N}_{k1}(\mathbf{x}_i), \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

where $\mathcal{N}_p(\mathbf{x}_i)$ denotes the set of p nearest neighbors of $\mathbf{x}_i$.

• Distant repulsion. This property enforces dissimilar data pairs in the original space to be far apart in the embedded space. Here we use the identical formulation with the 'local affinity' property as

$$\min_{\mathbf{Z}} \frac{1}{2} \sum_{i,j=1}^{n} s_{ij} ||\mathbf{z}_i - \mathbf{z}_j||^2 = \mathrm{Tr}(\mathbf{Z}\mathbf{L}_2\mathbf{Z}^T), \tag{6}$$

where $\mathbf{L}_2$ has similar property as $\mathbf{L}_1$ and

$$s_{ij} = \begin{cases} ||\mathbf{x}_i - \mathbf{x}_j||^2 \exp(-||\mathbf{x}_i - \mathbf{x}_j||^2/2t^2), & \text{if } \mathbf{x}_i \in \mathcal{N}_{k2}(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \mathcal{N}_{k2}(\mathbf{x}_i), \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Integrating Eqs.(4), (6) into the low-rank representation model, we can get the SPLRR model as follows:

$$\min_{\mathbf{Z},\mathbf{E}} ||\mathbf{Z}||_\star + \lambda||\mathbf{E}||_1 + \alpha\mathrm{Tr}(\mathbf{Z}\mathbf{L}_1\mathbf{Z}^T) + \beta\mathrm{Tr}(\mathbf{Z}\mathbf{L}_2\mathbf{Z}^T) \;\; s.t. \;\; \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}. \tag{8}$$

3.2 Solution to SPLRR

Similar to [8], Eq.(8) can be transformed into the following equivalent problem by introducing the auxiliary variable $\mathbf{J}$:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{E}} &||\mathbf{J}||_\star + \lambda||\mathbf{E}||_1 + \alpha\mathrm{Tr}(\mathbf{Z}\mathbf{L}_1\mathbf{Z}^T) + \beta\mathrm{Tr}(\mathbf{Z}\mathbf{L}_2\mathbf{Z}^T) \\ &s.t. \;\; \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E} \quad \text{and} \quad \mathbf{Z} = \mathbf{J}. \end{aligned} \tag{9}$$

In order to efficiently solve the optimization problem (9), the ALM method [7] is utilized. Thus, the Augmented Lagrange function w.r.t. (9) is:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{E},\mathbf{J},\mathbf{Y}_1,\mathbf{Y}_2} &||\mathbf{J}||_\star + \lambda||\mathbf{E}||_1 + \alpha\mathrm{Tr}(\mathbf{J}\mathbf{L}_1\mathbf{J}^T) + \beta\mathrm{Tr}(\mathbf{J}\mathbf{L}_2\mathbf{J}^T) + \langle\mathbf{Y}_1, \mathbf{X} - \mathbf{X}\mathbf{Z} - \mathbf{E}\rangle \\ &+\langle\mathbf{Y}_2, \mathbf{Z} - \mathbf{J}\rangle + \mu/2\left(||\mathbf{X} - \mathbf{X}\mathbf{Z} - \mathbf{E}||_F^2 + ||\mathbf{Z} - \mathbf{J}||_F^2\right). \end{aligned} \tag{10}$$

Obviously, we need to optimize this problem over one variable with others fixed. The subproblem w.r.t. each variable is convex and thus can provide corresponding unique optimal solutions. The optimization method to SPLRR is summarized in Algorithm 1. Note that, we give the relaxation of the objective when updating variable $\mathbf{J}$ and the derivation of Eq.(13) is as follows:

$$\begin{aligned} \mathcal{L} &= \min ||\mathbf{J}||_\star + \lambda||\mathbf{E}||_1 + \mathrm{Tr}\left(\mathbf{J}(\alpha\mathbf{L}_1 + \beta\mathbf{L}_2)\mathbf{J}^T\right) \\ &\leq \min ||\mathbf{J}||_\star + \lambda||\mathbf{E}||_1 + ||\mathbf{J}||_F \cdot ||\alpha\mathbf{L}_1 + \beta\mathbf{L}_2||_F \cdot ||\mathbf{J}||_F \\ &= \min ||\mathbf{J}||_\star + \lambda||\mathbf{E}||_1 + a\langle\mathbf{J}, \mathbf{J}\rangle \;\; (a \triangleq ||\alpha\mathbf{L}_1 + \beta\mathbf{L}_2||_F). \end{aligned} \tag{11}$$

For updating $\mathbf{J}$ while other variables fixed, we have

$$\begin{aligned} \mathbf{J} &= \arg\min ||\mathbf{J}||_\star + a\langle\mathbf{J}, \mathbf{J}\rangle - \langle\mathbf{Y}_2 + \mu\mathbf{Z}, \mathbf{J}\rangle + \mu/2\langle\mathbf{J}, \mathbf{J}\rangle \\ &= \arg\min ||\mathbf{J}||_\star + (a + \mu/2)\langle\mathbf{J}, \mathbf{J}\rangle - \mu\langle\mathbf{Z} + \mathbf{Y}_2/\mu, \mathbf{J}\rangle \\ &= \arg\min \frac{1}{2a+\mu}||\mathbf{J}||_\star + \frac{1}{2}||\mathbf{J} - \frac{\mu}{2a+\mu}\left(\mathbf{Z} + \frac{\mathbf{Y}_2}{\mu}\right)||_F^2. \end{aligned} \tag{12}$$

The SPLRR-based graph construction model is concluded in Algorithm 2.

Algorithm 1. Solving Problem (8) via Inexact ALM

Input: Data matrix $\mathbf{X}$; regularization parameters λ, α and β; parameters for constructing the affinity graph Laplacian and repulsion graph Laplacian.

Output: The affinity matrix $\mathbf{Z}$.

Initialization: set $\mathbf{Z} = \mathbf{J} = \mathbf{0}$, $\mathbf{E} = 0$, $\mathbf{Y}_1 = \mathbf{Y}_2 = \mathbf{0}$, $\mu = 10^{-6}$, $M = 10^{10}$, $\rho = 1.1$ and $\varepsilon = 10^{-8}$.

Construct the graph Laplacian $\mathbf{L}_1$ for local affinity and $\mathbf{L}_2$ for distant repulsion.

Repeat until converge:

- Updating $\mathbf{J}$:

$$\mathbf{J} = \arg\min \frac{1}{2a+\mu}||\mathbf{J}||_\star + \frac{1}{2}||\mathbf{J} - \frac{\mu}{2a+\mu}\left(\mathbf{Z} + \frac{\mathbf{Y}_2}{\mu}\right)||_F^2,\ a = ||\alpha\mathbf{L}_1 + \beta\mathbf{L}_2||_F \tag{13}$$

- Updating $\mathbf{Z}$: $\mathbf{Z} = (\mathbf{I} + \mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X} - \mathbf{X}^T\mathbf{E} + \mathbf{J} + (\mathbf{X}^T\mathbf{Y}_1 - \mathbf{Y}_2)/\mu)$
- Updating $\mathbf{E}$: $\mathbf{E} = \arg\min \frac{\lambda}{\mu}||\mathbf{E}||_1 + \frac{1}{2}||\mathbf{E} - (\mathbf{X} - \mathbf{X}\mathbf{Z} + \mathbf{Y}_1/\mu)||_F^2$
- Updating multipliers: $\mathbf{Y}_1 = \mathbf{Y}_1 + \mu(\mathbf{X} - \mathbf{X}\mathbf{Z} - \mathbf{E}),\quad \mathbf{Y}_2 = \mathbf{Y}_2 + \mu(\mathbf{Z} - \mathbf{J})$
- Updating the parameter μ by $\mu = \min(\rho\mu, M)$
- Checking the convergence conditions

$$||\mathbf{X} - \mathbf{X}\mathbf{Z} - \mathbf{E}||_\infty < \varepsilon \ \text{ and } \ ||\mathbf{Z} - \mathbf{J}||_\infty < \varepsilon$$

Algorithm 2. Graph construction based on SPLRR

Input: Data marix $\mathbf{X}$, parameters for computing affinity graph Laplacian and repulsion graph Laplacian.

Output: The weight matrix of SPLRR based graph $\mathbf{Z}$.

Normalize all the samples $\mathbf{x}_i$ to ℓ^2 unit norm.

Solve problem (8) using Algorithm 1. and get the optimal solution $\mathbf{Z}^\star$.

Normalize each column of $\mathbf{Z}^\star$ via $\mathbf{z}_i^\star = \mathbf{z}_i^\star/||\mathbf{z}_i^\star||_2$ and shrink entries in $\mathbf{Z}^\star$ by θ.

Construct the graph weight matrix $\mathbf{W}$ by $\mathbf{W} = \left(|\mathbf{Z}^\star| + (|\mathbf{Z}^\star|)^T\right)/2$.

4 Semi-supervised Classification

Denote $\mathbf{Y} = [(\mathbf{y}^1)^T; (\mathbf{y}^2)^T; \cdots; (\mathbf{y}^n)^T] \in \mathbb{R}^{n\times c}$ as the initial label matrix. If $\mathbf{x}_i$ is the unlabeled data, then $\mathbf{y}^i = \mathbf{0}$. If $\mathbf{x}_i$ is labeled data in class k, then the k-th element of $\mathbf{y}^i$ is 1 and the other elements of $\mathbf{y}^i$ are 0. Generally, graph based semi-supervised learning models solve the following problem [13]:

$$\min_{\mathbf{Q}} \mathrm{Tr}(\mathbf{Q}^T\tilde{\mathbf{L}}\mathbf{Q}) + \mathrm{Tr}((\mathbf{Q} - \mathbf{Y})^T\mathbf{U}(\mathbf{Q} - \mathbf{Y})), \tag{14}$$

where $\tilde{\mathbf{L}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$ is the normalized graph Laplacian, $\mathbf{U}$ is a diagonal matrix with the i-th diagonal element to control the impact of the initial label $\mathbf{y}^i$ of $\mathbf{x}_i$, $\mathbf{Q} \in \mathbb{R}^{n\times c}$ is the label matrix to be solved. For fair comparison, we simply set $U_{ii} = 1$ for all algorithms in our experiments.

Taking the derivative of Eq.(14) w.r.t $\mathbf{Q}$ and setting it to zero, we have:

$$\mathbf{L}\mathbf{Q} + \mathbf{U}(\mathbf{Q} - \mathbf{Y}) = \mathbf{0} \Rightarrow \mathbf{Q} = (\mathbf{L} + \mathbf{U})^{-1}(\mathbf{U}\mathbf{Y}). \tag{15}$$

5 Experiments

ORL, Extended Yale B and CMU PIE data sets are used in our experiments.

• **ORL**: There are 10 gray scale images for each of the 40 subjects. They were taken at different times, varying the lighting, facial expressions and facial details. Each face image is cropped and resized to 32×32.
• **Extended Yale B**: This database has 38 individuals, each subject having around 64 near frontal images under different illuminations. We simply use the first 50 cropped images of the first 20 individuals, and then resize them to 32×32.
• **PIE**: It contains 41368 images of 68 subjects with different poses, illumination and expressions. We only use their images in five near frontal poses and under different illuminations and expressions. The first 50 images of the first 20 subjects are selected. Each image is manually cropped and resized to size 32×32.

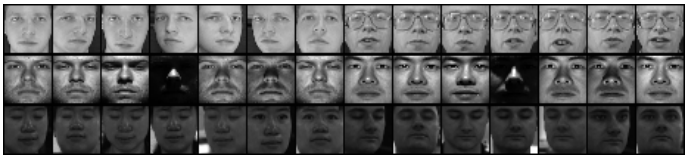

Fig. 1. Sample images from ORL, Extended Yale B and CMU PIE data sets

Some sample images from these three face databases are shown in Figure 1.

For evaluating the performance of proposed model, we compare SPLRR with some state-of-the-art graph construction models listed as follows:

• $k$nn-graph: The number of nearest neighbors for KNN1 and KNN2 are 4 and 8 respectively. The distance is measured using heat kernel and the kernel parameter is the average of squared Euclidean distances for all edged pairs.
• LNP [11]: We follow the pipeline of linear label propagation in to construct the graph. The neighborhood size in LNP is set to 40 to achieve the best results.
• ℓ^1 graph [5]: The ℓ^1 regularization item λ is empirically set to 0.01. And the ℓ^1 regularized least square problem is solved by $l1$-ls package.
• SPG graph [6]: we implement the SPG algorithm by setting n_{knn} as 10% of the size of data set and $\lambda = 0.001$.
• LRR graph [12]: the λ in LRR is set to an near optimal value 0.1.
• SPLRR graph: The number of nearest neighbors for 'local affinity' and 'distant repulsion' constraints are 4 and 20 respectively. The kernel parameter is empirically set as $0.1 \times \sqrt{-\bar{d}/\ln\{0.1/k\}}$ [9] ($\bar{d}$ is the average of squared Euclidean distances for all edged pairs on the graph, k is the neighbor number to construct the neighborhood graph). The hyper-parameters λ, α and β are empirically set as 0.5, 0.9 and 0.1 for all data sets.

For each face data set, we randomly select 10% to 60% face images per subject as labeled samples and the rest as unlabeled samples. Tables 1 reports the face recognition results on these three data sets.For each configuration, we conduct 50

Table 1. Experimental results on ORL, Extended Yale B and PIE(mean±std-dev%)

ORL	KNN1	KNN2	LNP	L1-Graph	SPG	LRR	SPLRR
10%	64.98±2.07	53.47±2.80	71.21±2.30	61.89±2.69	65.92±2.42	71.63±2.45	**77.31±2.17**
20%	74.07±2.71	63.97±2.69	82.01±2.20	76.16±2.73	78.74±1.98	83.67±2.12	**87.73±2.17**
30%	79.42±2.24	68.93±2.72	88.06±2.45	84.61±2.32	86.24±2.36	88.66±1.79	**91.66±1.67**
40%	81.16±2.43	71.64±2.82	91.43±1.68	89.31±1.90	90.72±1.64	91.35±1.72	**94.11±1.84**
50%	82.76±2.41	72.83±2.44	93.15±1.74	92.47±1.69	93.20±1.64	93.54±1.41	**95.93±1.51**
60%	83.23±2.20	74.75±2.69	94.69±1.64	94.25±1.59	95.46±1.50	94.46±1.55	**96.94±1.38**
YaleB	KNN1	KNN2	LNP	L1-Graph	SPG	LRR	SPLRR
10%	73.04±1.62	55.82±2.91	86.22±1.52	77.69±1.74	82.92±1.62	87.18±1.35	**92.96±0.98**
20%	77.13±1.31	63.03±2.39	90.86±1.14	87.58±1.05	89.84±1.16	92.36±1.17	**95.86±0.92**
30%	80.12±1.38	67.29±1.91	92.45±0.72	92.13±1.15	92.75±0.99	93.98±0.99	**97.31±0.65**
40%	81.49±1.35	69.86±2.46	93.42±0.86	94.39±0.94	94.26±0.77	95.42±0.81	**98.23±0.53**
50%	83.50±1.43	72.24±2.45	93.85±0.82	95.90±1.03	95.59±0.72	96.13±0.82	**98.75±0.42**
60%	84.25±2.03	74.74±2.42	94.89±0.98	97.29±0.93	96.37±0.85	96.73±0.86	**99.12±0.45**
PIE	KNN1	KNN2	LNP	L1-Graph	SPG	LRR	SPLRR
10%	49.38±2.83	40.09±2.11	67.50±2.77	63.60±2.35	65.29±2.16	75.93±2.05	**77.07±2.32**
20%	59.22±2.24	51.99±1.93	79.11±1.47	76.43±1.18	77.80±1.70	87.00±1.42	**88.11±1.65**
30%	64.69±1.73	60.76±1.77	83.54±1.73	82.93±1.44	83.82±1.26	90.33±1.17	**91.20±1.23**
40%	67.12±1.91	68.56±1.45	87.02±1.35	86.99±1.25	87.23±1.25	92.39±1.05	**93.56±1.09**
50%	69.87±2.09	75.11±1.10	89.08±1.38	89.34±1.02	89.35±1.41	93.79±1.11	**94.60±1.04**
60%	71.49±2.04	81.15±1.11	90.07±1.53	90.96±1.32	91.60±1.63	94.79±1.02	**95.47±1.34**

independent runs for each algorithm. The mean accuracy as well as the standard deviation of the performance are reported.

From the experimental results, we can observe that: 1) The LRR based methods, both LRR and SPLRR, perform consistently well on three data sets, which suggests that the LRR induced affinity matrix is efficient for semi-supervised learning. Generally, LNP, L1-Graph and SPG have similar performances while the baseline KNN graph has the lowest accuracy. 2) SPLRR graph consistently achieves the lowest classification error rates over other graphs even with low labeling percentages which can be observed from results on ORL and Extended Yale B data sets. When there are only 10% labeled samples, SPLRR can still obtain very high accuracies. This means the structure information of data set, which should be preserved in transformation process, is important when building an informative graph for semi-supervised learning.

6 Conclusion

In this paper, we have proposed a new graph construction model for semi-supervised face recognition, called structural preserving low-rank representation (SPLRR). SPLRR constructs the graph with preserving the structure of data set, which enforces the local affinity property to be preserved without distorting of the distant repulsion property. As a result, the proposed model derives

a informative graph and shows the best performance in the comparison with state-of-the-art methods for semi-supervised face recognition.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (Grant No.61272248), the National Basic Research Program of China (Grant No.2013CB329401) and the Science and Technology Commission of Shanghai Municipality (Grant No.13511500200).

References

1. Cai, D., He, X., Han, J., Huang, T.S.: Graph regularized nonnegative matrix factorization for data representation. IEEE TPAMI 33(8), 1548–1560 (2011)
2. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optimiz. 20(4), 1956–1982 (2010)
3. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? J. ACM 58(3) (2011)
4. Carreira-Perpinán, M.A.: The elastic embedding algorithm for dimensionality reduction. In: ICML, pp. 167–174 (2010)
5. Cheng, B., Yang, J., Yan, S., Fu, Y., Huang, T.S.: Learning with ℓ^1-graph for image analysis. IEEE TIP 19(4), 858–866 (2010)
6. He, R., Zheng, W.S., Hu, B.G., Kong, X.W.: Nonnegative sparse coding for discriminative semi-supervised learning. In: CVPR, pp. 2849–2856 (2011)
7. Lin, Z., Chen, M., Ma, Y.: The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint arXiv:1009.5055 (2010)
8. Liu, G., Lin, Z., Yu, Y.: Robust subspace segmentation by low-rank representation. In: ICML, vol. 3 (2010)
9. Nie, F., Xiang, S., Jia, Y., Zhang, C.: Semi-supervised orthogonal discriminant analysis via label propagation. Pattern Recogn. 42(11), 2615–2627 (2009)
10. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319–2323 (2000)
11. Wang, F., Zhang, C.: Label propagation through linear neighborhoods. IEEE TKDE 20(1), 55–67 (2008)
12. Yang, S., Wang, X., Wang, M., Han, Y., Jiao, L.: Semi-supervised low-rank representation graph for pattern recognition. IET IP 7(2), 131–136 (2013)
13. Zhou, D., Bousquet, O., Lal, T.N., Weston, J., Schölkopf, B.: Learning with local and global consistency. In: NIPS, vol. 16 (2004)