International Journal of Physical Sciences Vol. 7(10), pp. 1633 - 1640, 2 March, 2012
Available online at http://www.academicjournals.org/IJPS
DOI: 10.5897/IJPS12.008
ISSN 1992 - 1950 ©2012 Academic Journals

Full Length Research Paper

Learning with power l_1-graph for single labeled image biometric recognition

Fei Zang, Jiang-She Zhang and Chun-Xia Zhang

School of Science and State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China.

Accepted 07 February, 2012

Single labeled biometric recognition is one of the main challenges to graph-based transductive classification learning. To enhance the recognition rate of single labeled problem, sparse representation provides a feasible strategy for representation learning. In this paper, we developed a power l_1-graph learning technique for semi-supervised learning, called label propagation by power l_1-graph (LPPG). Different from all existing graph-based methods, we assume that the similarity relationship in the label space is a power function in the sample space. What is important is that the determinated power value measured by sparseness is given. Our method characterizes the second sparse processing, and seeks to find a reasonable label propagation way. This characteristic makes our algorithm more intuitive and more powerful than those methods based on the original l_1-graph. This proposed method is applied to biometrics recognition and the experiment results show that our algorithm consistently outperforms those original l_1-graph-based methods. This demonstrates that our method is a good choice for real-world biometrics applications, especially when there is only one labeled image.

Key words: Power function, sparse reconstruction, label propagation, biometrics recognition.

INTRODUCTION

In many practical applications such as law enforcement, driver license or passport card identification, only one labeled sample per person is available. Under such scenario, most of semi-supervised classification techniques perform badly, even fail to work. For example, when the number of labeled samples from each class is very small, the estimated values of statistical measures, such as probability density function or class conditional entropy, are very gross, which makes those methods depending on statistical measures of labeled samples achieve undesirable results. Therefore, special tricks are generally required to deal with the single labeled sample problem.

Recent graph-based semi-supervised learning representative of liner neighborhood propagation (LNP) (Wang et al., 2006; Wang et al., 2009) provides a feasible strategy to deal with such problem. The basic idea of LNP which can be cast into a second-order intrinsic Gaussian Markov random field framework is to predict the label of a sample according to its neighbors in a linear way. Different from many approaches, LNP provides a graph structure construction method by introducing multiple-wise edges instead of pairwise edges, and presents an effective scheme to estimate the weights for such multiple-wise edges. The key of LNP is to construct a connected graph by neighborhood information of each sample and estimate the weight at the cost of minimizing linear reconstruction error. The k nearest neighbor approach used in LNP is easier to obtain a connected graph, while the defined neighbors usually cannot well characterize the real geometrical relations among samples. Note that the adjacency structure of the graph is already fixed during the first step and the consequent graph weight calculation step will be constrained by these neighborhood relations. Meanwhile, that distance measured by the usual Euclidean norm is sensitive to noise. To solve those problems of k nearest graph, l_1-graph is proposed in (Cheng et al., 2010; Yan and Wang, 2010; Qiao et al., 2010). This graph is constructed by sparse representation,

*Corresponding author. E-mail: jszhang@mail.xjtu.edu.cn. Tel: +86-29-82665961.

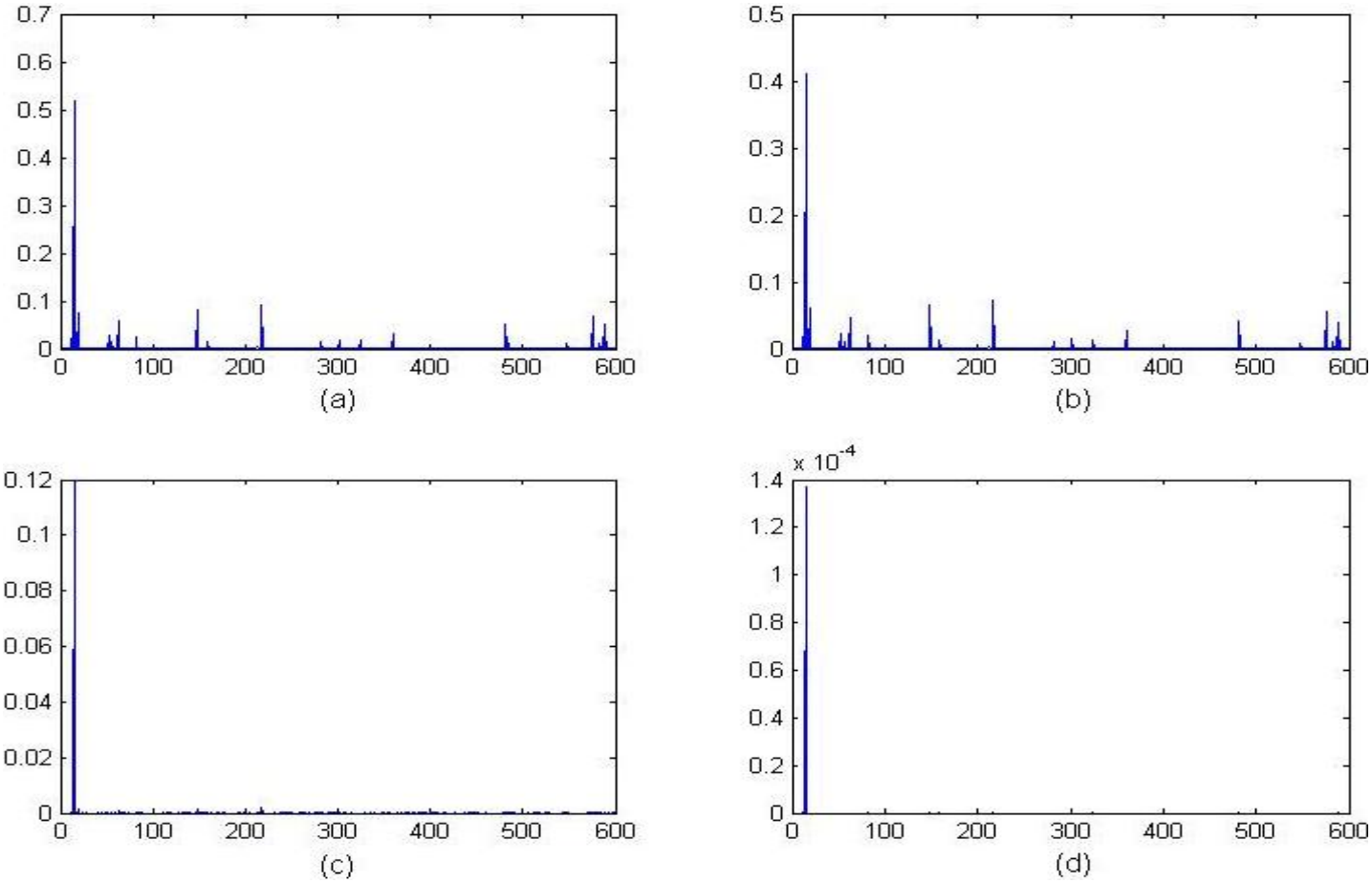

Figure 1. Sparse reconstruction coefficients for different methods are shown on PolyU palmprint database. (a) original l_1-graph, (b) normalized l_1-graph which the sum of all construction coefficients is one, (c) power l_1-graph, where power parameter α is given by LPPG and (d) power l_1-graph, where power parameter α is fixed as 10.

and thus, the local structure in data is automatically modeled instead of being manually predefined. That is to say, it avoids the difficulty of parameter selection as in LNP. Moreover, the graph adjacency structure and graph weights are determined simultaneously by solving the l_1-norm optimization process. And then l_1-graph builds the direct relationship between the graph structure and the edge weight.

However, there are some shortcomings in the l_1-graph. Figure 1(a) displays the sparse reconstruction coefficients by original l_1-graph. Figure 1(b) shows the normalized coefficients which appeared in Figure 1(a). It is noted that both Figure 1(a) and (b) trail a long tail. That is to say, the coefficients of those samples which come from different classes from the represented sample are non-zeros. Although, the role of each tail is tiny, the cumulative effect of those tails could not be ignored. So the reconstructed label of the represented sample is made to contain much non-positive information which is a disadvantage to correctly predict label.

Motivated by the aforementioned observations, we propose a new graph-based transductive classification method called label propagation by power l_1-graph (LPPG), for single labeled image biometric recognition. We first make a bold assumption that the similarity relationship in the label space is a power function in the sample space. From the point of view of sparseness, our method can be regarded as a second sparse processing. It is important that we give the determined relationship between the sparseness and power parameter. Figure 1(c) lists the sparse reconstruction coefficients by our LPPG. Obviously, the trail phenomenon is eliminated. That is to say, we intuitively emphasize the role of samples which have the same class with the represented sample in label propagation processing. A mass of experimental results on biometric recognition demonstrate the reasonability of that assumption.

Label propagation by power l_1-graph (LPPG)

Give a sample set $\boldsymbol{X}=[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n] \in R^{m\times n}$ and a label set $L=\{1, 2, \cdots, c\}$. The first l samples $\boldsymbol{X}_l=[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_l]$ have labels $\{y_1, y_2, \cdots, y_c\}$ and the remaining samples $\boldsymbol{X}_u=[\boldsymbol{x}_{l+1}, \boldsymbol{x}_{l+2}, \cdots, \boldsymbol{x}_n]$ are unlabeled. In this paper, we assume that each class has only one labeled sample.

Power l_1-graph

Suppose we have an underdetermined system of linear equations, $\boldsymbol{x}_i=\boldsymbol{X}_i\boldsymbol{w}_i$, where $\boldsymbol{x}_i$ is the ith sample from the sample set $\boldsymbol{x}$, $\boldsymbol{w}_i$ is the vector of the sparse reconstruction coefficients, and $\boldsymbol{X}_i=[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{i-1}, \boldsymbol{x}_{i+1}, \cdots, \boldsymbol{x}_n] \in R^{m\times(n-1)}$ is the overcomplete dictionary. Sparsity motivates us to seek the sparsest solution by solving the following l_1-norm optimization problem:

$$\min_{\boldsymbol{w}_i} \|\boldsymbol{w}_i\|_1 \quad s.t. \begin{cases} \boldsymbol{x}_i=\boldsymbol{X}_i\boldsymbol{w}_i \\ \boldsymbol{w}_i \geq 0 \\ \boldsymbol{e}_i^T\boldsymbol{w}_i=1 \end{cases} \tag{1}$$

where $\|\cdot\|_1$ denotes the l_1-norm which sums the absolute value of each entries in a vector, and $\boldsymbol{e}_i=[1, 1, \cdots, 1] \in R^{n-1}$. And then the sparse reconstruction relationship matrix $\boldsymbol{W}$ can be defined as:

$$\begin{cases} \boldsymbol{W}_{ij}=\boldsymbol{w}_i^j, & i>j \\ \boldsymbol{W}_{ij}=\boldsymbol{w}_i^{j-1}, & i<j \\ \boldsymbol{W}_{ij}=0, & i=j \end{cases} \tag{2}$$

where $\boldsymbol{w}_i^j$ denotes the jth entry of $\boldsymbol{w}_i$. To achieve our purpose, we assume the relationship matrix $\boldsymbol{S}$ in label space is a power function of $\boldsymbol{W}$ which reflects the sparse reconstructive relationship in sample space, and we have

$$\boldsymbol{S}=\boldsymbol{W}^{\alpha} \tag{3}$$

where $\alpha \geq 1$ is a power parameter. Our power l_1-graph is degraded to original l_1-graph when α is fixed as 1. Because $0 \leq \boldsymbol{w}_i^j \leq 1$, if $\alpha \to \infty$ and then $\boldsymbol{S} \to 0$. Figure 1(c) demonstrates that case which almost eliminates the role of all samples, and then, how to give a power value becomes a key problem in LPPG.

Label propagation

In our LPPG algorithm, which assumes that, for the ith sample $\boldsymbol{x}_i$, the label of $\boldsymbol{x}_i$ can be linearly reconstructed by other samples' labels. Here, we estimate the label of all samples by minimizing

$$\min \sum_{i=1}^{n} \left\| \boldsymbol{y}_i - \sum_{j=1}^{n} \boldsymbol{S}_{ij}\boldsymbol{y}_j \right\|_2^2 \tag{4}$$

where $\boldsymbol{y}_i$ denotes label vector of the sample $\boldsymbol{x}_i$. Based on basic algebraic knowledge, it can be easily inferred that

$$\begin{aligned} &\sum_{i=1}^{n} \left\| \boldsymbol{y}_i - \sum_{j=1}^{n} \boldsymbol{S}_{ij}\boldsymbol{y}_j \right\|_2^2 \\ &= \|\boldsymbol{Y}(\boldsymbol{I}-\boldsymbol{S})\|_2^2 \\ &= trace(\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{S})^T(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{Y}) \\ &= trace(\boldsymbol{Y}^T\boldsymbol{M}\boldsymbol{Y}) \end{aligned} \tag{5}$$

where $\boldsymbol{I}$ is an identity matrix and $\boldsymbol{M}=(\boldsymbol{I}-\boldsymbol{S})^T(\boldsymbol{I}-\boldsymbol{S})$ is a symmetric matrix, $trace(\cdot)$ denotes the trace of matrix. For the general semi-supervised learning, we let $\boldsymbol{Y}=[\boldsymbol{Y}_l; \boldsymbol{Y}_u]$ denote the label matrix of all samples, where $\boldsymbol{Y}_l$ denotes the label submatrix of the labeled samples and $\boldsymbol{Y}_u$ denotes the initial label submatrix of the unlabeled samples. We divide the matrix $\boldsymbol{M}$ into four parts:

$$\boldsymbol{M}=\begin{pmatrix} \boldsymbol{M}_{ll} & \boldsymbol{M}_{lu} \\ \boldsymbol{M}_{ul} & \boldsymbol{M}_{uu} \end{pmatrix} \tag{6}$$

And then take derivative of $trace(\boldsymbol{Y}^T\boldsymbol{M}\boldsymbol{Y})$ with respect to $\boldsymbol{Y}$, we have

$$\boldsymbol{M}\boldsymbol{Y}=0 \tag{7}$$

And that is

$$\begin{cases} \boldsymbol{M}_{ll}\boldsymbol{Y}_l+\boldsymbol{M}_{lu}\boldsymbol{Y}_u=0 \\ \boldsymbol{M}_{ul}\boldsymbol{Y}_l+\boldsymbol{M}_{uu}\boldsymbol{Y}_u=0 \end{cases} \tag{8}$$

Finally, we get the matrix form relation between the labeled and unlabeled samples as

$$\boldsymbol{Y}_u = -\boldsymbol{M}_{uu}^{-1}\boldsymbol{M}_{ul}\boldsymbol{Y}_l \quad (9)$$

Power parameter

Here, according to the sparseness of all samples, we give a suitable value of power parameter α. The sparseness of vector is the number of nonzero in that vector. Let p_i denote the sparseness of $\boldsymbol{w}_i$, that is, $p_i = \|\boldsymbol{w}_i\|_0$. p_{aver} indicates the average sparseness, $p_{aver} = \frac{1}{n}\sum_{i=1}^{n} p_i$, and $p_{\min}$ is the minimization value of all p_i. And then, the power parameter α can be determined by

$$\alpha = \log(p_{aver} + p_{\min} - 1) \quad (10)$$

Sparse coding

Here, three different sparse coding methods used in our paper are briefly reviewed. Suppose we have an underdetermined system of linear equations: $\boldsymbol{z} = \boldsymbol{X}\boldsymbol{\beta}$, where $\boldsymbol{Z}$ is the sample to be represented, $\boldsymbol{\beta}$ is the vector of the reconstruction coefficients, and $\boldsymbol{X}$ is the overcomplete dictionary. In practice, the sample $\boldsymbol{Z}$ and the basis dictionary $\boldsymbol{X}$ are known, and representation coefficient $\boldsymbol{\beta}$ is to be computed according to the different representation demands. Sparsity motivates us to seek the sparsest solution to $\boldsymbol{z} = \boldsymbol{X}\boldsymbol{\beta}$ by solving the following optimization problem:

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_0, \ s.t. \quad \boldsymbol{z} = \boldsymbol{X}\boldsymbol{\beta}$$

where $\|\cdot\|_0$ denotes the l_0 norm that counts the number of nonzero entries in a vector. However, the problem of finding the sparsest solution to an underdetermined system of linear equations is NP-hard problem. Recent development in the emerging theory of sparse representation and compressed sensing reveals that, if the solution $\boldsymbol{\beta}$ sought is sparse enough, the solution of the l_0-minimization problem is equal to the solution to the following l_1-minimization problem

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1, \ s.t. \quad \boldsymbol{z} = \boldsymbol{X}\boldsymbol{\beta} \quad (11)$$

where $\|\cdot\|_1$ denotes the l_1 norm which sums the absolute value of each entries in a vector. Here, we denote the algorithm using Equation 11 to construct a graph by l_1-graph0. This graph weights are computed via an active set algorithm based on (Lee et al., 2006).

To deal with occlusions and corruptions, Wright et al. (2009) further proposed a robust linear model as $\boldsymbol{z} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$, where $\boldsymbol{e} \in R^m$ is a noise item of errors. Assuming that the noise item $\boldsymbol{e}$ also has a sparse representation, the l_1-graph can compute a robust weight vector $\boldsymbol{\beta}$ as follows

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1 + \|\boldsymbol{e}\|_1 \quad s.t. \quad \|\boldsymbol{z} - (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e})\|_2 \le \eta \quad (12)$$

We denote the algorithm using Equation 12 to construct a graph by l_1-graph 1. The graph weights are computed by Equation 12 via a primal-dual algorithm for linear programming based on (Boyd and Vandenberghe, 2004). However, in many applications, the noise level η is unknown beforehand (Elhamifar and Vidal, 2009). In such cases, the Lasso optimization algorithm (Tibshirani, 1996) can be used to recover the sparse solution from

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{z} - (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e})\|_2^2 + \lambda(\|\boldsymbol{\beta}\|_1 + \|\boldsymbol{e}\|_1) \quad (13)$$

where λ can be viewed as a regularizer constant. We denote the algorithm using Equation 13 to construct a graph by l_1-graph2. This graph weights are computed via an active set algorithm based on Lee et al. (2006). The regularized item λ is empirically to 0.05.

EXPERIMENTS

Data sets preparation

To demonstrate the effectiveness of our algorithm, we gathered almost all the popular face databases for our experiments. Seven face databases are used, including AR, facial recognition technology (FERET), Georgia Tech (GT), INDIAN, Japanese female facial expression (JAFFE), ORL and CMU PIE databases. For all of these databases, facial images are aligned by fixing the locations of two eyes. Meanwhile, one palmprint data set is used to test our algorithm.

(1) The ORL database (Olivetti and Oracle Research Laboratory, 1994) contains 40 distinct subjects, each of which has 10 different images. For some subjects, the images are taken at different times, with the varied lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). Each image is manually cropped and normalized to the size of 32×32 pixels. Figure 2 shows some face images.

(2) The GT database (Georgia, 2007) consists of 50 subjects with 15 images per subject. It characterizes several variations such as pose, expression, cluttered background and illumination. That is to say,

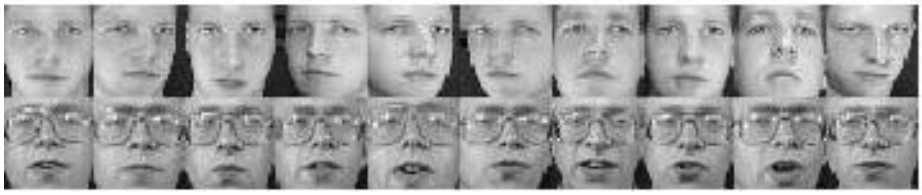

Figure 2. Samples of the cropped images in the ORL database.

Figure 3. Samples of the cropped images in the GT database.

Figure 4. Samples of the cropped images in a subset of the FERET database.

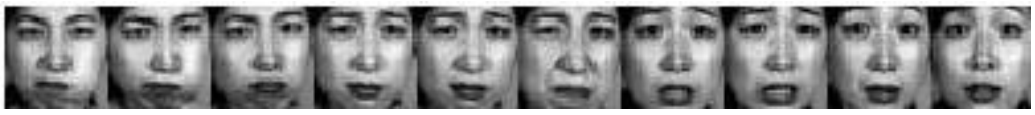

Figure 5. Samples of the cropped images in the JAFFE database.

Figure 6. Samples of the cropped images in a subset of the AR database.

these images have large variations in both pose and expression and some illumination changes. Images are converted to gray scale and cropped into the size of 32×32. Some face images of two people are shown in Figure 3.

(3) The FERET face image database (Phillips et al., 2000) 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 32×32 pixels and further preprocessed by histogram equalization. Figure 4 shows some of the faces with varying expressions, illumination and poses.

(4) The JAFFE database (Lyons et al., 1999) contains 213 images of 7 facial expressions (6 basic facial expressions and 1 neutral) posed by 10 Japanese female models. All the cropped images are normalized to the 32×32 pixels with 256 gray level per pixel. Figure 5 depicts some face images from this database.

(5) The AR database (Martinez and Benavente, 1998) consists of over 4000 face images of 126 individuals. For each individual, 26 images are taken in two sessions and each section contains 13 images. These images include front view of faces with different expressions, illuminations and occlusions. In our experiments, we use a subset of the AR face database. This subset contains 1400 images corresponding to 100 people, where each person has 14 different images with the changed illumination and expressions. All the images were resized to 48×48. Figure 6 shows some sample face images.

Figure 7. Samples of the cropped images in the INDIAN database.

Figure 8. From the top to the bottom, sample images of one person respectively come from C09 and C29 in the CMU PIE database.

(6) The INDIAN database (Jain and Mukherjee, 2002) contains images of 40 distinct subjects with eleven different poses for each individual. All images have a bright homogeneous background and the subjects are in an upright, frontal position. For each individual, images include the following pose for the face: looking front, looking left, looking right, looking up, looking up towards left, looking up towards right, looking down. In addition to the variation in pose, images with four emotions - neutral, smile, laughter, sad/disgust - are also included for each individual. In this paper, we only use those twenty-one female's face images. Each image is manually cropped and normalized to the size of 32-by-32 pixels. As an example, some images are shown in Figure 7.
(7) The CMU PIE database (Sim et al., 2003) includes 41,368 images of 68 people, each person with 13 different poses, 43 different illumination conditions and with 4 different expressions. In our experiments, two subset of near frontal poses (C09 and C29) are used. Each subset has 24 different pictures of 68 people, and each picture is manually cropped and normalized to the size of 48×48 pixels. Figure 8 shows the sample cropped face images from the chosen database.

We also use a general palmprint object recognition data set, that is, the PolyU Palmprint database (Zhang et al., 2003) contains 7752 grayscale images corresponding to 386 different palms. Around twenty samples from each of these palms were collected in two sessions, where around 10 samples were captured in the first session and the second session respectively. We only use the first 30 different palms and each picture is manually cropped and normalized to the size of 64-by-64 pixels.

TRANSDUCTIVE CLASSIFICATION

Here, we test our approach on several single labeled image biometric recognition problems. We compare LPPG algorithm with original l_1-graph and LNP, where power parameter is determined by cross-validation (LPPG(cv)) and Equation 10. As a baseline, we also give the classification results of 1-NN classifier directly using the raw data. In our experiments, we kept 98% information in the sense of reconstruction error in the principal components analysis (PCA) projection process. For each database, we average the results over 20 random splits.
The average recognition rates (%) and the standard deviations are shown in Tables 1 and 2. From the Tables 1 and 2, we can see that the results of LNP depend on the size of neighborhood k which is difficulty to be determined. Those methods based on spare reconstruction are superior to LNP. That demonstrates spare reconstruction and reflects more intrinsic similarity relationship between samples than liner reconstruction. What is important is that, LPPG consistently outperforms those methods based on the original l_1-graph. This shows that those samples have more sparse structure and our assumption reflects more nature characteristic than original l_1-graph. Figure 9 demonstrates the relationship of the recognition rates of LPPG versus the value of power parameter α. From Figure 9, we can see that the recognition rates of LPPG get higher with the increasing the value of power parameter α. And then, the recognition rates of LPPG get lower when the value of power parameter α is greater than some value. This phenomenon shows the information which is captured by LPPG is varied with the change of α. So, how to determine the suitable value of power parameter α is very important. It is noted that the results of LPPG(cv) are only a little better than LPPG, which demonstrates that the power parameter selection in Equation 10 is reasonable. However, the computation cost of LPPG(cv) is much larger than LPPG.

CONCLUSION

To deal with single labeled biometric recognition problem, in this letter, we have presented a transductive classification approach by constructing a power l_1-graph. Different from all graph-based semi-supervised learning methods, our algorithm assumes that the sample space and the label space share the different reconstruction weight. Furthermore, we measure the power parameter in LPPG by the sparseness of all samples. Extensive numerical experiments confirmed that our adventurous

Table 1. Recognition rates (%) of different methods on the four databases with different configurations. Note that LPPG(cv) denotes that the power parameter in LPPG is determined by 10-fold cross-validation.

Dataset	N/A	FERET	INDIAN	JAFFE	PolyU
Baseline	N/A	39.84±2.28	50.43±3.40	69.58±6.44	66.47±2.60
LNP	$k=5$	59.86±2.33	61.84±3.87	82.54±3.64	73.48±16.38
	$k=8$	61.31±2.62	62.51±2.41	81.95±6.34	79.28±2.79
l_1-graph	l_1-magic	68.24±1.84	57.21±3.03	71.36±4.62	68.46±2.76
	lasso1	66.17±1.49	60.97±3.19	72.70±5.43	83.66±1.85
	lasso2	66.16±1.45	61.80±2.89	74.46±5.58	85.85±2.07
LPPG(cv)	l_1-magic	72.33±1.21	64.09±2.49	79.13±4.28	93.30±1.59
	lasso1	70.89±1.28	66.19±2.80	79.72±5.69	93.69±1.90
	lasso2	70.69±1.20	66.26±2.80	80.89±5.83	94.73±1.86
LPPG	l_1-magic	70.51±1.65	62.36±2.86	76.08±5.02	91.10±2.57
	lasso1	69.90±1.46	64.74±2.82	78.29±5.63	92.33±2.06
	lasso2	69.86±1.34	64.35±2.94	79.32±6.08	93.80±2.13

Table 2. Recognition rates (%) of different methods on the five databases with different configurations. Note that LPPG(cv) denotes that the power parameter in LPPG is determined by 10-fold cross-validation.

Dataset	N/A	AR	GT	ORL	C09	C29
Baseline	N/A	21.84±1.21	44.39±2.20	55.07±2.61	18.04±1.05	19.48±0.98
LNP	$k=5$	30.83±1.56	58.79±2.03	73.11±2.32	35.26±1.38	37.64±1.57
	$k=8$	31.53±1.37	58.94±2.20	73.06±1.91	35.40±1.50	39.39±1.58
l_1-graph	l_1-magic	72.12±1.78	44.78±1.95	70.34±2.20	76.72±2.70	78.04±2.63
	lasso1	57.02±1.74	49.28±2.28	75.14±2.32	73.40±2.74	69.12±2.73
	lasso2	58.44±1.82	49.71±2.04	61.80±2.89	73.34±2.71	69.48±2.75
LPPG(cv)	l_1-magic	79.94±1.80	50.65±2.37	78.14±2.20	84.28±2.47	81.41±2.27
	lasso1	63.11±1.89	55.66±1.96	80.41±1.73	81.91±3.02	79.69±2.76
	lasso2	64.39±1.74	55.89±1.80	80.78±1.75	81.84±3.03	79.67±2.71
LPPG	l_1-magic	78.85±1.85	49.49±2.46	76.85±2.24	83.58±2.66	80.78±2.27
	lasso1	62.28±1.95	54.91±2.04	79.21±1.88	81.59±3.06	78.90±2.57
	lasso2	63.66±1.68	55.15±1.98	79.75±1.81	81.55±3.06	78.59±2.47

assumption is very reasonable. In future work, we will demonstrate the capability of LPPG to deal with other tasks, as text classification or image segmentation and apply our idea to solve the single labeled problem in feature extraction and clustering learning.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the anonymous referees as well as the Editor and Associate Editor for their valuable comments which lead to

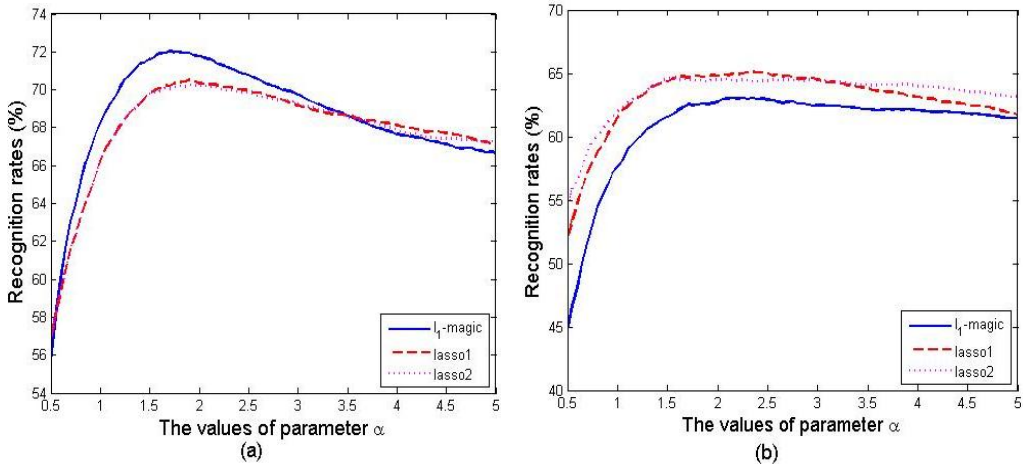

Figure 9. Recognition rates (%) of LPPG under three different sparse coding methods versus the variation of power parameter α on different face databases. (a) FERET, (b) INDIAN.

substantial improvements of the paper. This work was supported by the National Natural Science Foundation of China (Grant Nos. 61075006) and the Research Fund for the Doctoral Program of Higher Education of China (No. 20100201120048).

REFERENCES

Boyd S, Vandenberghe L (2004). Convex optimization. Cambridge University Press.

Cheng B, Yang J, Yan SC, Fun Y, Huang TS (2010). Learing with l_1-graph for image analysis. IEEE Trans. Image Process., 19(4): 858-866.

Elhamifar E, Vidal R (2009). Sparse subspace clustering. Proceeding of the 2009 IEEE CVPR.

Georgia Tech. Face Database (2007). http://www.anefian.com/face_reco.htm.

Jain V, Mukherjee A (2002). The Indian face database. http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/.

Lee H, Battle A, Raina R, Ng AY (2006). Efficient sparse coding algorithms. In Neural Information Processing Systems (NIPS).

Lyons MJ, Budynek J, Akamatsu S (1999). Automatic classification of single facial images. IEEE Trans. Patt. Anal. Machine Intell., 21(12): 1357-1362.

Martinez AM, Benavente R (1998). The AR face database. CVC Technical Report #24. Rep. 24.

Olivetti & Oracle Research Laboratory (1994). The Olivetti & Oracle Research Laboratory Face Database of Faces, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

Phillips PJ, Moon H, Rizvi SA, Rauss PJ (2000). The FERET evaluation methodology for face recognition algorithms. IEEE Trans. Patt. Anal. Machine Intell., 22(10): 1090-1104.

Qiao LS, Chen SC, Tan XY (2010). Sparsity preserving projections with applications to face recognition. Patt. Recognit., 43(1): 331-341.

Sim T, Baker S, Bsat M (2003). The CMU pose, illuminlation, and expression database. IEEE Trans. Patt. Anal. Machine Intell., 25(12): 1615-1618.

Tibshirani R (1996). Regression shrinkage and selection via the lasso. J. the Royal Statistical Soc. B., 58(1): 267-288.

Wang F, Wang JD, Zhang CS, Shen HC (2006). Supervised classification using linear neighborhood propagation. Processing of the 2006 IEEE CVPR., pp. 160-167.

Wang JD, Wang F, Zhang CS, Shen HC, Quan L (2009). Linear neighborhood propagation and its applications. IEEE Trans. Patt. Anal. Machine Intell., 31(9): 1600-1615.

Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009). Robust face recognition via sparse representation. IEEE Trans. Patt. Anal. Machine Intell., 31(2): 210-227.

Yan SC, Wang H (2010). Semi-supervised learning by sparse representation. Accepted by SIAM.

Zhang D, Kong WK, You J, Wong M (2003). On-line palmprint identification. IEEE Trans. Patt. Anal. Machine Intell., 25(9): 1041-1050.