LOCALITY PRESERVING KSVD FOR NONLINEAR MANIFOLD LEARNING

Yin Zhou, Jinglun Gao, Kenneth E. Barner

University of Delaware, Newark, DE, USA, 19716
{zhouyin, gao, barner}@udel.edu

ABSTRACT

Discovering the intrinsic low-dimensional structure from high-dimensional observation space (*e.g.,* images, videos), in many cases, is critical to successful recognition. However, many existing nonlinear manifold learning (NML) algorithms have quadratic or cubic complexity in the number of data, which makes these algorithms computationally exorbitant in processing real-world large-scale datasets. Randomly selecting a subset of data points is very likely to place NML algorithms at the risk of local optima, leading to poor performance. This paper proposes a novel algorithm called Locality Preserving KSVD (LP-KSVD), which can effectively learn a small number of dictionary atoms as locality-preserving landmark points on the nonlinear manifold. Based on the atoms, the computational complexity of NML algorithms can be greatly reduced while the low-dimensional embedding quality is improved. Experimental results show that LP-KSVD successfully preserves the geometric structure of various nonlinear manifolds and it outperforms state-of-the-art dictionary learning algorithms (MOD, K-SVD and LLC) in our preliminary study on face recognition.

Index Terms— Dimensionality reduction, Manifold learning, Dictionary learning, Sparse coding, Face recognition

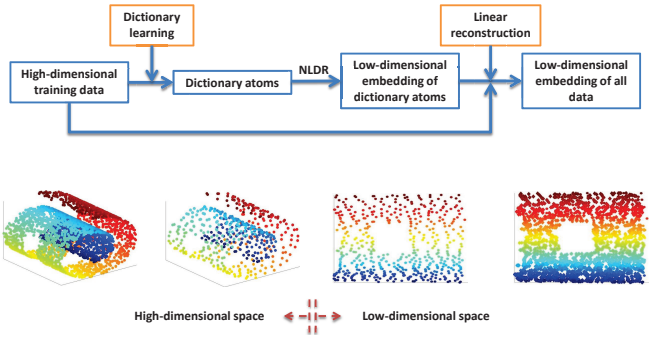

Fig. 1: Overview of the proposed method. Given training data in high-dimensional observation space, a representational and locality-preserving dictionary is learned. Then, the low-dimensional embedding of the atoms is computed via some NML algorithm. Finally, using the geometric relationships among training data and the atoms in observation space, the low-dimensional embedding of training data is reconstructed as linear combinations of the low-dimensional embedding of the atoms.

1. INTRODUCTION

One of the central tasks in signal analysis and pattern recognition is to seek effective representations for real-world high-dimensional data [1]. Dimensionality reduction is an important technique that discovers the most succinct and intrinsic forms of representation of the original high-dimensional data, allowing more effective learning and prediction. Linear dimensionality reduction algorithms *e.g.,* Principle Component Analysis (PCA) [2], Linear Discriminative Analysis (LDA) [3], have been widely applied in the past decades due to their simplicity and efficiency. Such algorithms, however, typically are not capable of exploiting the nonlinear structure of a data manifold and therefore are not suitable for processing complex datasets [4]. In recent years, nonlinear manifold learning (NML) algorithms have been developed to effectively discover the intrinsic low-dimensional structure of the nonlinear manifold. Examples are ISOMAP [5], Locally Linear Embedding (LLE) [6], Hessian LLE [7], Laplacian Eigenmap [1], Diffusion map [8], Local Tangent Space Alignment (LSTA) [9], etc.

However, many current NML algorithms are of quadratic or cubic complexity in the number of data, which diminishes the applicability of these algorithms in real-world large-scale datasets [10]. Efforts have been made on selecting a subset of training data as landmark points on the manifold to improve the efficiency of NML algorithms. Landmark points are meaningful points that preserve the local geometric structure of a manifold. Silva and Tenenbaum [11] suggested using a subset of randomly selected data points, which, however, is susceptible to local optimum leading to poor performance. The authors of [10] proposed to use LASSO (Least Absolute Shrinkage and Selection Operator) regression for selecting landmark points, which is of high computational cost due to the ℓ_1 minimization. Thus, it is still an open and challenging problem of learning landmark points effectively.

This paper proposes a novel algorithm–Locality Preserving KSVD (LP-KSVD)–to learn a compact, representational, and locality-preserving dictionary. The overview of the proposed method is illustrated in Fig. 1. From the training set, a small number of atoms are learned as landmark points. NML algorithms can very efficiently compute the low-dimensional embedding of these landmark atoms. Then, based on the atoms, the original high-dimensional manifold as well as the low-dimensional embedding can be accurately reconstructed via locally linear representation [12]. Learning a dictionary of landmark atoms has the advantages of robust to outliers, random initialization, imbalanced data distribution and allowing better generalization capability.

Our work is different from sparse coding *e.g.,* K-SVD [13], the method of optimal directions (MOD) [14], since 1) the employed local coding has closed-form solution and is more efficient compared to sparsity driven algorithms, *e.g.,* Orthogonal Matching Pursuit (OMP) [15]; 2) the dictionary is optimized for both its capability in representation and its locality preservation capability, which is in contrast to sparse coding which only solves for a representational dictionary. Experimental results support the promising performance of LP-KSVD.

2. RELATED WORK

Sparse coding expresses an input signal succinctly by only a few atoms from a dictionary or codebook. This model has been successfully applied to a variety of problems in image processing and computer vision such as image restoration [16–18], image denoising [13, 14, 19], image classification [20–22], etc. In particular, Aharon *et al* [13] developed K-SVD to efficiently learn an overcomplete dictionary from training data and achieved impressive results on image denoising and image compression. K-SVD only focuses on optimizing a representational dictionary without considering the locality-preserving capability of the dictionary. Another similar algorithm is the method of optimal directions (MOD) [14]. Moreover, sparse coding algorithms are typically of high computational complexity due to the stage of solving sparse coefficients.

Locality-based coding is recently developed [23, 24]. Specifically, Yu *et al.* [23] introduced Locally Coordinate Coding (LCC) to approximate nonlinear functions as a linear combination of anchor points and experimentally showed that locality can be more essential than sparsity in representing data distributed on nonlinear manifold. Unfortunately, their coding strategy is based on a modification of LASSO and hence is still computationally expensive. For the purpose of efficient learning, Wang *et al.* [24] further proposed Locality-constrained Linear Coding (LLC) as a fast approximation to LCC and achieved state-of-the-art results on image classification. Nevertheless, little efforts has been made on learning a representational and locality-preserving dictionary to improve the performance of NML algorithms.

3. PROPOSED METHOD

3.1. Problem Formulation

Let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in \mathbb{R}^{m \times N}$ be the training set, which contains m-dimensional N samples. Suppose $\mathbf{y}_i$ reside on a smooth manifold of intrinsic dimension n ($n \ll m$), which is embedded into $\mathbb{R}^m$. The goal here is the joint achievement of two objectives. First is establishing a compact dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K] \in \mathbb{R}^{m \times K}$ such that linear combinations of $\mathbf{d}_i$ can approximate the nonlinear manifold $\mathcal{M} \subset \mathbb{R}^m$. As we have no access to the true $\mathcal{M}$, $\mathbf{D}$ is estimated based on $\mathbf{Y}$. The second objective is learning $\mathbf{d}_i$ as landmark points, which are capable of preserving the locality on $\mathcal{M}$. The dictionary learning problem (LP-KSVD) is thus formalized as:

$$< \mathbf{D}, \mathbf{X} >= \arg\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \tag{1}$$
$$\text{s.t.} \begin{cases} x_{ij} = 0 & \text{if } \mathbf{d}_i \notin \Omega_\tau(\mathbf{y}_j) \quad \forall i, j \\ \mathbf{1}^{\mathrm{T}}\mathbf{x}_j = 1 & \forall j \end{cases}$$

where the reconstruction error term measures the fitness of $\mathbf{D}$ to $\mathbf{Y}$; the matrix $\mathbf{X} \in \mathbb{R}^{K \times N}$ contains N local reconstruction codes, with $\mathbf{x}_j$ being the code for reconstructing $\mathbf{y}_j$; and $\Omega_\tau(\mathbf{y}_j)$ denotes the neighborhood containing τ nearest dictionary atoms of $\mathbf{y}_j$ in terms of Euclidean distance. We call the first constraint local cardinality constraint, which requires that every training sample can only be reconstructed by its τ nearest-neighbor dictionary atoms. The second constraint allows the reconstruction coefficients to be invariant to translation of the data [6]. In order to achieve faithful reconstruction, Yu *et al.* [23] suggested that $\mathbf{d}_i$ be sufficiently close to $\mathcal{M}$. Moreover, to learn $\mathbf{d}_i$ as landmark points, we further require each $\mathbf{d}_i$ to be locally representational with respect to a small patch on $\mathcal{M}$. We will show that satisfying this requirement, the proposed LP-KSVD algorithm can effectively achieve the aforementioned two objectives.

3.2. Optimization

The proposed LP-KSVD solves Eq. (1) iteratively by alternating between the two variables, *i.e.,* we first fix $\mathbf{D}$ and solve for the best coefficient matrix $\mathbf{X}$ and then, we update $\mathbf{D}$ as well as $\mathbf{X}$ [13] jointly. The iterations are terminated if either the objective function value is below some preset threshold or a maximum number of iterations has been reached.

3.2.1. Solving for local reconstruction codes

Fixing $\mathbf{D}$ which is initialized or learned from previous iteration, the $\mathbf{X}$ defined in Eq. (1) can be obtained by equivalently solving [6]:

$$\min_{\mathbf{x}_j \in \mathbf{X}} \left\| \mathbf{y}_j - \sum_{i=1}^{K} x_{ij}\mathbf{d}_i \right\|_2^2 \tag{2}$$
$$\text{s.t.} \begin{cases} x_{ij} = 0 & \text{if } \mathbf{d}_i \notin \Omega_\tau(\mathbf{y}_j) \quad \forall i \\ \mathbf{1}^{\mathrm{T}}\mathbf{x}_j = 1 & \forall j \end{cases}$$

where $\mathbf{x}_j$ represents the j-th column in $\mathbf{X}$, containing linear representation coefficients for reconstructing $\mathbf{y}_j$, and x_{ij} is the i-th element in $\mathbf{x}_j$. Let $\hat{\mathbf{x}}_j$ be a subvector containing only the nonzero elements in $\mathbf{x}_j$. The closed-form solution to Eq. (2) is given as [6]:

$$\hat{\mathbf{x}}_j = (\mathbf{G} + \eta\mathbf{I})\backslash\mathbf{1} \tag{3}$$

and

$$\hat{\mathbf{x}}_j = \hat{\mathbf{x}}_j / \sum \hat{x}_{ij}, \tag{4}$$

where $\mathbf{G} = (\Omega_\tau - \mathbf{y}_j\mathbf{1}^{\mathrm{T}})^{\mathrm{T}}(\Omega_\tau - \mathbf{y}_j\mathbf{1}^{\mathrm{T}})$ is the local covariance matrix, η is a small constant to secure numerical stability and the operator $\backslash$ means matrix inversion[1]. Here, we write $\Omega_\tau(\mathbf{y}_i)$ as Ω_τ for simplicity.

3.2.2. Local Dictionary Optimization

In this step, we continue to minimize the objective function by updating $\mathbf{D}$ and $\mathbf{X}$ jointly. We optimize each dictionary atom individually. Denoting the k-th atom in $\mathbf{D}$ as $\mathbf{d}_k \in \mathbb{R}^m$ and the k-th row of $\mathbf{X}$ as $\mathbf{x}_{k*} \in \mathbb{R}^{1 \times N}$, we update $\mathbf{d}_k$ as follows. First, let $\mathbf{E}_k = \mathbf{Y} - \sum_{t \neq k} \mathbf{d}_t\mathbf{x}_{t*}$; next, minimize $\|\mathbf{E}_k - \mathbf{d}_k\mathbf{x}_{k*}\|_F^2$ with respect to $\mathbf{d}_k$ and $\mathbf{x}_{k*}$. As $\|\mathbf{E}_k - \mathbf{d}_k\mathbf{x}_{k*}\|_F^2$ is only affected by the $\omega = \|\mathbf{x}_{k*}\|_0$ nonzero entries in $\mathbf{x}_{k*}$, it can be equivalently minimized by solving:

$$< \mathbf{d}_k, \tilde{\mathbf{x}}_{k*} >= \arg\min_{\mathbf{d}_k, \tilde{\mathbf{x}}_{k*}} \|\tilde{\mathbf{E}}_k - \mathbf{d}_k\tilde{\mathbf{x}}_{k*}\|_F^2, \tag{5}$$

where $\tilde{\mathbf{E}}_k$ contains the ω error columns from $\mathbf{E}_k$, which only associates with the nonzero entries in $\mathbf{x}_{k*}$, *i.e.,* the succinct row vector $\tilde{\mathbf{x}}_{k*} \in \mathbb{R}^{1 \times \omega}$. Actually, minimizing Eq.(5) yields simultaneous update of $\mathbf{d}_k$ and $\tilde{\mathbf{x}}_{k*}$ but only $\mathbf{d}_k$ is of our interest as the output of the algorithm.

KSVD [13] simply treats Eq.(5) as a rank-1 matrix approximation problem, and solves the best shape for $\mathbf{d}_k$. This, however, only yields a representational $\mathbf{D}$, and has no guarantee of preserving the local geometric structure of $\mathcal{M}$. In order to learn atoms as landmark points, we enforce each atom to be locally representational with respect to a small patch on $\mathcal{M}$. Recall that the local cardinality

[1] We have followed the same notation as [6, 24]. For detailed derivation of Eq. (3)(4), please refer to [6] or http://www.cs.nyu.edu/ roweis/lle/.

constraint requires each $\mathbf{y}_j$ to be reconstructed by its nearest dictionary atoms. We therefore define Λ as a small patch on $\mathcal{M}$, which is a neighborhood set containing the ω training samples that are concurrently choosing $\mathbf{d}_k$ as one of their nearest neighbors. Thus, by jointly achieving the minimization of Eq.(5) and the best local representation with respect to Λ, we are able to learn a representational and locality-preserving $\mathbf{D}$.

Proposition 1 (Local Dictionary Optimization)**.** *Let $\Lambda \in \mathbb{R}^{m\times\omega}$ be a small neighborhood containing ω training samples that concurrently select $\mathbf{d}_k \in \mathbb{R}^m$ as a nearest neighbor. Let $\tilde{\mathbf{E}}_k \in \mathbb{R}^{m\times\omega}$ be the error matrix corresponding to Λ. Define the local representation error (LRE) as $\sum_{\mathbf{y}_i\in\Lambda} \|\mathbf{d}_k - \mathbf{y}_i\|_2^2$. Then, the $\mathbf{d}_k^{new}$ and the $\tilde{\mathbf{x}}_{k*}^{new}$ that minimize $\|\tilde{\mathbf{E}}_k - \mathbf{d}_k\tilde{\mathbf{x}}_{k*}\|_F^2$ and yield the minimum LRE are given as:*

$$\begin{aligned} \mathbf{U}\Delta\mathbf{V}^{\mathrm{T}} &= \tilde{\mathbf{E}}_k \\ \mathbf{d}_k^{new} &= s\mathbf{u} \qquad (6) \\ \tilde{\mathbf{x}}_{k*}^{new} &= \frac{\Delta(1,1)\mathbf{v}^{\mathrm{T}}}{s} \qquad (7) \\ s &= \frac{1}{\omega}\sum_{\mathbf{y}_i\in\Lambda}\frac{\mathbf{u}^{\mathrm{T}}}{\|\mathbf{u}\|_2}\mathbf{y}_i \qquad (8) \end{aligned}$$

where $\mathbf{u}$ and $\mathbf{v}$ are the first columns of $\mathbf{U}$ and $\mathbf{V}$; s is the gain factor (scale) of $\mathbf{d}_k^{new}$.

Proof. We adopt a sequential optimization strategy. On minimizing $\|\tilde{\mathbf{E}}_k - \mathbf{d}_k\tilde{\mathbf{x}}_{k*}\|_F^2$, applying SVD yields an optimal rank-1 matrix approximation solution as

$$\begin{aligned} \mathbf{U}\Delta\mathbf{V}^{\mathrm{T}} &= \tilde{\mathbf{E}}_k \\ \mathbf{d}_k^{new} &= \mathbf{u} \qquad (9) \\ \tilde{\mathbf{x}}_{k*}^{new} &= \Delta(1,1)\mathbf{v}^{\mathrm{T}} \qquad (10) \end{aligned}$$

where $\mathbf{u}$ and $\mathbf{v}$ are the first column of $\mathbf{U}$ and $\mathbf{V}$. Note that $\mathbf{u}$ is a unit vector indicating the shape (direction) of $\mathbf{d}_k^{new}$. Then with $\mathbf{u}$ fixed, the gain factor s of $\mathbf{d}_k^{new}$ that minimizes LRE can be obtained by solving

$$\begin{aligned} \min_{s} \quad & \sum_i \|(s\tfrac{\mathbf{u}}{\|\mathbf{u}\|_2} - \mathbf{y}_i)\|_2^2 \\ \text{s.t.} \quad & \mathbf{y}_i \in \Lambda \end{aligned} \qquad (11)$$

and the optimal s is given by

$$s = \frac{1}{\omega}\sum_{\mathbf{y}_i\in\Lambda}\frac{\mathbf{u}^{\mathrm{T}}}{\|\mathbf{u}\|_2}\mathbf{y}_i. \qquad (12)$$

Without affecting the closest rank-1 matrix approximation to $\tilde{\mathbf{E}}_k$, simultaneously multiplying the right hand side (RHS) of Eq. (9) and dividing the RHS of Eq. (10) by s yields the desired updates. □

As LP-KSVD essentially minimizes reconstruction error in the same way as K-SVD [13], convergence to a local minimum is guaranteed.

3.3. Test Sample Embedding

Having obtained $\mathbf{D} \in \mathbb{R}^{m\times K}$, its low-dimensional embedding $\mathbf{B} \in \mathbb{R}^{n\times K}$, where $n < m$, can be computed by a particular NML algorithm. Given a test sample $\mathbf{z} \in \mathbb{R}^m$, its low-dimensional embedding $\mathbf{h} \in \mathbb{R}^n$ can be linearly reconstructed as $\mathbf{h} = \mathbf{B}\mathbf{x}$, where $\mathbf{x}$ is the local reconstruction code of $\mathbf{z}$ over $\mathbf{D}$, as recommended in [6].

4. EXPERIMENTAL RESULTS

In this section, we first demonstrate the capability of the proposed LP-KSVD in preserving the local geometric structure of manifolds, by visualizing the dimensionality reduction results on several synthetic datasets. Then, the performance of LP-KSVD is further verified over two popular face recognition databases, *i.e.,* Extended YaleB Database [25] and CMU PIE Database [26]. The parameter τ is set to 2 uniformly.

4.1. Visualization on Synthetic Datasets

The 3-dimensional synthetic datasets[2] employed are shown in Fig.2. To demonstrate the wide applicability of LP-KSVD to various NML algorithms, Hessian LLE [7], Laplacian Eigenmap [1] and LLE [6] are employed to Swiss Hole, noisy Toroidal Helix, and Gaussian, with the number of nearest neighbors $k = 6, 3, 8$ respectively.

In Fig.2, from the top row to the bottom row are the visualizations (3D manifold and 2D embeddings) of 3000 training data, randomly selected K training samples, K dictionary atoms learnt by LP-KSVD, and 2000 test data. Over all these datasets, the dictionary atoms are learned as landmark points equidistributed on the original 3D manifold and preserve well the local geometric structure, *e.g.,* the non-convex region (hole) in Fig.2(a), the ring-shape and connectivity in Fig.2(b) and the curvature shape in Fig.2(c). Moreover, the smooth color displayed by the 2D embeddings verifies that test samples are successfully mapped into the subspace spanned by the dictionary atoms. In contrast, as shown in the 2nd row of Fig.2, NML algorithms have difficulty in computing low-dimensional embeddings over a small number of randomly selected training samples, due to the nonuniform distribution of data points.

4.2. Face Recognition

The Extended YaleB Database contains 2414 frontal face images of 38 subjects [25]. For each subject, we randomly select half of the images (about 32 per person) for training and the other half for testing. As in [27], we use a subset of the CMU PIE Database [26], *i.e.*, C05, C07, C09, C27, and C29, in which images are nearly frontal poses and are taken under varying conditions of illumination and expression. The subset yields a total number of 11554 images with about 170 images per subject. Following [27], a random selection of 130 images per person are employed to form the training set, and the rest of the database are for testing. For both databases, the images are normalized to 32×32 pixels and are preprocessed by histogram equalization.

We evaluate LP-KSVD and compare it with MOD [14], K-SVD [13] and the recently proposed LLC [24] by classifying test images embedded in the low-dimensional subspace spanned by dictionary atoms. For all dictionary learning (DL) algorithms, a structured dictionary is learned as $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_C]$, where $\mathbf{D}_i$ is the sub-dictionary for class i. We initialize all DL algorithms with 8 atoms per subject, which yields a dictionary of 304 atoms for Extended YaleB Database and a dictionary of 544 atoms for CMU PIE Database. Linear and nonlinear dimensionality reduction are performed by PCA and LLE respectively. The nearest neighbor

[2] Accessible at:http://www.math.ucla.edu/∼wittman/mani/

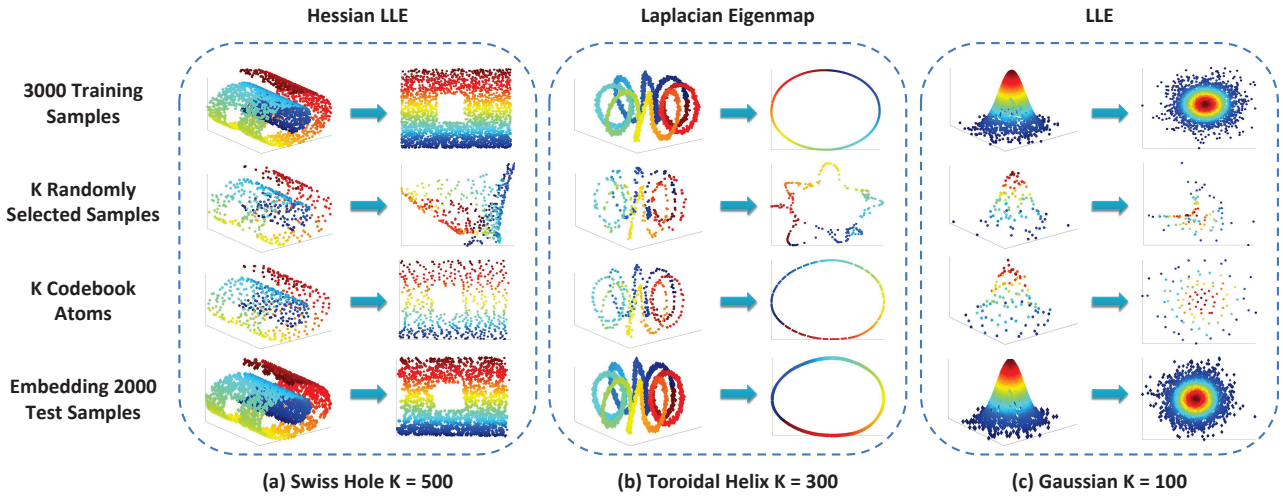

Fig. 2: Visualizing dimensionality reduction results on synthetic datasets.

(NN) classifier is employed. We report recognition rate as the average over 30 repetitions. The sparsity factor of OMP [15] is set to 16 for MOD and K-SVD, as recommended in [22]. Increasing the sparsity factor can lead to a marginal accuracy improvement but will cause significant higher computation time [28].

We choose All Train as the baseline method, which represents the results obtained by performing NML on the entire training set. Random stands for the result obtained by employing randomly selected training samples as the dictionary. Recognition performances of all approaches are illustrated in Fig. 3. We can see that among three, *i.e.,* Fig. 3(a), 3(b) and 3(d), out of the four evaluation scenarios, LP-KSVD outperforms other methods in recognition rate. For example, as shown in Fig. 3(a), LP-KSVD achieves the highest recognition rate 95.7% and leads K-SVD and All Train in the second place by 1.2%. In Fig. 3(c), different DL algorithms display diverse performances and LP-KSVD yields the same highest accuracy 96.7% as that of All Train. However, considering the fact that the compression rate ($\frac{\sharp\text{atoms}}{\sharp\text{training samples}}$) for CMU PIE Database is only 6.2%, the significant saving in computation complexity makes LP-KSVD a very competitive approach. Note that in Fig. 3(b) and 3(d), all methods achieve similar accuracies. This is due to the fact that PCA seeks the optimal projection directions that best preserve the data variance and is insensitive to the nonlinear structure of a manifold. We also evaluate the efficiency of LP-KSVD and compare it with other methods. The training time for all methods include the computation time of both dictionary learning (DL) and training data embedding (TrDE). From Table 1, we can see that over large-scale databases (*e.g.,* CMU PIE), the efficiency of LP-KSVD is up to 149.3 times higher than that of All Train, which further validates the usability of the proposed method.

5. CONCLUSION AND FUTURE WORK

We propose a novel LP-KSVD algorithm to learn dictionary atoms as landmark points, which can preserve the locality on nonlinear manifold. Experimental results validate that the proposed method is superior to existing DL algorithms in terms of greatly reducing computational complexity while yielding higher classification rate. Future work includes refining the local reconstruction coding strategy by incorporating local sparse coding. In addition, LP-KSVD should be further evaluated over more datasets.

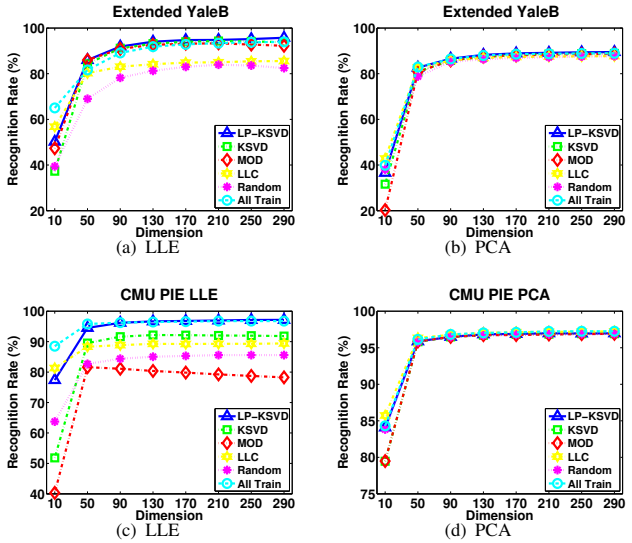

Fig. 3: Classification results over two face databases. Results in (a) and (c) are obtained via LLE dimensionality reduction; results in (b) and (d) are obtained via PCA dimensionality reduction. Note: the parameter k of LLE is set to 60 for both the Extended YaleB and CMU PIE databases. All Train represents the results based on performing NML algorithms on the entire training set.

Table 1: Overall execution time (seconds) over two face databases. For all methods, the overall time is the sum of dictionary learning time and training data embedding time.

	Extended YaleB		CMU PIE	
	Overall Time	Speedup	Overall Time	Speedup
All Train	22.1577 s	No	11807.3121 s	No
K-SVD [13]	59.4014 s	No	848.3236 s	13.9x
MOD [14]	41.8138 s	No	611.1109 s	19.3x
LLC [24]	11.6593 s	1.9x	69.2321 s	170.5x
LP-KSVD	10.0008 s	2.2x	79.0830 s	149.3x

6. ACKNOWLEDGEMENT

The authors would like to thank NSF Grant No. 0812458 for funding and thank the reviewers for their helpful comments.

7. REFERENCES

[1] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, June 2003.

[2] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.

[3] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[4] A. Geiger, R. Urtasun, and T. Darrell, "Rank priors for continuous non-linear dimensionality reduction," in *CVPR 2009. IEEE Conference on*, 2009, pp. 880 –887.

[5] J. B. Tenenbaum, V.de Silva, and J. C. Landford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319 – 2323, 2000.

[6] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323 – 2326, 2000.

[7] D. L. Donoho and C. Grimes, "Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data," 2003.

[8] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, "Diffusion maps, spectral clustering and reaction coordinates of dynamical systems," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 113 – 127, 2006.

[9] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," *SIAM J. Sci. Comput.*, vol. 26, pp. 313–338, January 2005.

[10] J. Silva, J. Marques, and J. Lemos, "Selecting landmark points for sparse manifold learning," in *Advances in Neural Information Processing Systems 18*, pp. 1241–1248. 2006.

[11] V. de Silva and J. B. Tenenbaum, "Global versus local methods in nonlinear dimensionality reduction," in *Advances in Neural Information Processing Systems 15*. 2002.

[12] L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds," *J. Mach. Learn. Res.*, vol. 4, pp. 119–155, Dec. 2003.

[13] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionarys for sparse representation," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4311 – 4322, 11 2006.

[14] K. Engan, S.O. Aase, and J.H. Husoy, "Frame based signal compression using method of optimal directions (mod)," in *ISCAS '99.*

[15] J.A. Tropp and A.C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Trans.*, vol. 53, no. 12, pp. 4655 –4666, 2007.

[16] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE TIP*, pp. 53 –69, 2008.

[17] H. Zhang, J. Yang, Y. Zhang, N.M. Nasrabadi, and T.S. Huang, "Close the loop: Joint blind image restoration and recognition with sparse representation prior," in *ICCV 2011*, pp. 770 –777.

[18] S. Xie and S. Rahardja, "An alternating direction method for frame-based image deblurring with balanced regularization," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 1061 – 1064.

[19] C. Studer and R.G. Baraniuk, "Dictionary learning from sparsely corrupted or compressed signals," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 3341 –3344.

[20] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. PAMI*, vol. 31, no. 2, pp. 210 – 227, 2 2009.

[21] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *CVPR 2010.*

[22] Q. Zhang and B. Li, "Discriminative k-svd for dictionary learning in face recognition," *CVPR 2010.*

[23] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *NIPS09*.

[24] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *CVPR 2010.*

[25] K.C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. PAMI*, vol. 27, no. 5, pp. 684–698, 2005.

[26] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression database," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 12, pp. 1615 – 1618, dec. 2003.

[27] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *ICCV 2007.*, 2007, pp. 1 –7.

[28] Y. Zhou and K. E. Barner, "Locality constrained dictionary learning for nonlinear dimensionality reduction," *Signal Processing Letters, IEEE*, vol. 20, no. 4, pp. 335–338, April.