[14] H. Liu and H. Motoda, *Instance Selection and Construction for Data Mining*. Norwell, MA: Kluwer, 2001.

[15] H. Fayed and A. Atiya, "A novel template reduction approach for the k-nearest neighbor method," *IEEE Trans. Neural Netw.*, vol. 20, no. 5, pp. 890–896, May 2009.

[16] H. A. Fayed, S. R. Hashem, and A. F. Atiya, "Self-generating prototypes for pattern classification," *Pattern Recognit.*, vol. 40, no. 5, pp. 1498–1509, May 2007.

[17] W. Lam, C.-K. Keung, and D. Liu, "Discovering useful concept prototypes for classification based on filtering and abstraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1075–1090, Aug. 2002.

[18] C.-L. Chang, "Finding prototypes for nearest neighbor classifiers," *IEEE Trans. Comput.*, vol. 23, no. 11, pp. 1179–1184, Nov. 1974.

[19] C. H. Chen and A. Jóźwik, "A sample set condensation algorithm for the class sensitive artificial neural network," *Pattern Recognit. Lett.*, vol. 17, no. 8, pp. 819–823, Jul. 1996.

[20] M. Lozano, J. M. Sotoca, J. S. Sánchez, F. Pla, E. Pekalska, and R. P. W. Duin, "Experimental study on prototype optimisation algorithms for prototype-based classification in vector spaces," *Pattern Recognit.*, vol. 39, no. 10, pp. 1827–1838, Oct. 2006.

[21] J. S. Sánchez, "High training set size reduction by space partitioning and prototype abstraction," *Pattern Recognit.*, vol. 37, no. 7, pp. 1561–1564, 2004.

[22] T. Kohonen, "The self organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.

[23] J. Li, M. T. Manry, C. Yu, and D. R. Wilson, "Prototype classifier design with pruning," *Int. J. Artif. Intell. Tools*, vol. 14, nos. 1–2, pp. 261–280, 2005.

[24] S.-W. Kim and B. J. Oommen, "Enhancing prototype reduction schemes with LVQ3-type algorithms," *Pattern Recognit.*, vol. 36, no. 5, pp. 1083–1093, May 2003.

[25] F. Fernández and P. Isasi, "Evolutionary design of nearest prototype classifiers," *J. Heuristics*, vol. 10, no. 4, pp. 431–454, Jul. 2004.

[26] L. Nanni and A. Lumini, "Particle swarm optimization for prototype reduction," *Neurocomputing*, vol. 72, nos. 4–6, pp. 1092–1097, Jan. 2009.

[27] A. Cervantes, I. M. Galván, and P. Isasi, "AMPSO: A new particle swarm method for nearest neighborhood classification," *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.*, vol. 39, no. 5, pp. 1082–1091, Oct. 2009.

[28] R. Storn and K. Price, "Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, Dec. 1997.

[29] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization* (Natural Computing Series), G. Rozenberg, T. Bäck, A. E. Eiben, J. N. Kok, and H. P. Spaink, Eds. New York: Springer-Verlag, 2005.

[30] K. V. Price, *An Introduction to Differential Evolution*. London, U.K.: McGraw-Hill, 1999.

[31] F. Neri and V. Tirronen, "Scale factor local search in differential evolution," *Memetic Comput.*, vol. 1, no. 2, pp. 153–171, 2009.

[32] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Mateo, CA: Morgan Kaufmann, 2005.

[33] J. Alcalá-Fdez, L. Sánchez, S. García, M. J. del Jesus, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, V. M. Rivas, J. C. Fernández, and F. Herrera, "KEEL: A software tool to assess evolutionary algorithms for data mining problems," *Soft Comput.*, vol. 13, no. 3, pp. 307–318, Oct. 2009.

[34] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, "KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," *J. Multiple-Valued Logic Soft Comput.*, 2010, to be published.

[35] S. García, A. Fernández, J. Luengo, and F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," *Inform. Sci.*, vol. 180, no. 10, pp. 2044–2064, May 2010.

[36] S. García, A. Fernández, J. Luengo, and F. Herrera, "A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability," *Soft Comput.*, vol. 13, no. 10, pp. 959–977, Apr. 2009.

Linear Discriminant Analysis for Signatures

Seungil Huh and Donghun Lee

***Abstract*— We propose signature linear discriminant analysis (signature-LDA) as an extension of LDA that can be applied to signatures, which are known to be more informative representations of local image features than vector representations, such as visual word histograms. Based on earth mover's distances between signatures, signature-LDA does not require vectorization of local image features in contrast to LDA, which is one of the main limitations of classical LDA. Therefore, signature-LDA minimizes the loss of intrinsic information of local image features while selecting more discriminating features using label information. Empirical evidence on texture databases shows that signature-LDA improves upon state-of-the-art approaches for texture image classification and outperforms other feature selection methods for local image features.**

***Index Terms*— Earth mover's distance, feature selection, linear discriminant analysis, signature, texture classification.**

I. Introduction

Linear discriminant analysis (LDA) is one of the most popular feature selection methods for classification tasks and has been widely used in machine learning and computer vision [1]–[3]. Numerous extensions of LDA have been exploited to improve the original LDA [4]–[11]. One of the main limitations of classical LDA is that it can be applied to data only in a vector space. Therefore, the given data, each of which is not represented as a vector, must be transformed into vectors before applying LDA.

Recently, local feature approaches have been widely used for texture image classification, object recognition, and image matching tasks because these approaches are descriptive as well as robust to occlusion and background clutter [12]–[14]. After obtaining local features through detection of the region of interest [12], [15]–[17] and descriptor computation [13], [18], visual word histograms comprising the frequencies of each visual word in images are typically employed to represent a large number of local features [13].

However, visual histograms often suffer from quantization or binning problems [19]. To resolve the problems, signatures [19] (see Section II for the definition) can be utilized to represent local image features instead of visual histograms. Previous works [18], [20] have shown empirically that signature-based methods generally outperform visual histograms on texture classification and object recognition tasks. Signatures are more informative than visual word histograms in that a signature is expressed not as a vector but as a collection of cluster center vectors and their relative weights. However, in contrast to histograms, signatures are beyond the scope

Manuscript received September 24, 2009; revised August 2, 2010 and October 19, 2010; accepted October 21, 2010. Date of publication November 11, 2010; date of current version November 30, 2010.

S. Huh is with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: seungilh@cs.cmu.edu).

D. Lee was with Carnegie Mellon University, Pittsburgh, PA 15213 USA. He is now with the Department of Computer Science, Princeton University, Princeton, NJ 08540 USA (e-mail: donghunl@princeton.edu).

Digital Object Identifier 10.1109/TNN.2010.2090047

of what classical LDA or existing LDA extensions can handle, which makes signatures-based approaches less competitive.

In this brief, we propose an extension of LDA, called signature-LDA, which can be applied to signatures. Since the optimal transformation matrix, which discriminates the different classes, is computed based on earth mover's distances (EMDs) [19] between signatures, vectorization of signatures is not required to be applied to signature-LDA. Therefore, signature-LDA minimizes the loss of intrinsic information of local image features while selecting more informative features with label information. The empirical evidence on texture databases shows that signature-LDA not only improves the performance of signature/EMD-based texture classification but also outperforms other feature selection methods applied to local image features.

Section II provides an overview of background knowledge for our signature-LDA method. Then, we formulate signature-LDA and describe its process and properties in Section III. The experimental results are reported in Section IV, followed by conclusions and future work in Section V.

II. Background

In this section, we describe the concept of signature and EMD [19]. In addition, we introduce the pairwise formulation of LDA and local Fisher discriminant analysis (LFDA) [8].

A signature is a set of clusters represented by their centers and relative weights, where the center of a cluster is the mean vector of samples assigned to the cluster, and its weight is the fraction of samples assigned to the cluster [19]. Thus, a signature X is formalized as

$$X = \{(c_1, w_1), (c_2, w_2), \ldots, (c_m, w_m)\} \tag{1}$$

where c_i and w_i denote the center and the weight of the ith cluster, respectively, and m denotes the number of clusters.

The EMD between two signatures is defined as the minimum cost of transforming one signature to another [19]. More formally, EMD is computed by solving the following optimization problem.

Let X_1 and X_2 be two signatures such that

$$\begin{aligned} X_1 &= \left\{\left(c_1^{(1)}, w_1^{(1)}\right), \left(c_2^{(1)}, w_2^{(1)}\right), \ldots, \left(c_{m_1}^{(1)}, w_{m_1}^{(1)}\right)\right\} \\ X_2 &= \left\{\left(c_1^{(2)}, w_1^{(2)}\right), \left(c_2^{(2)}, w_2^{(2)}\right), \ldots, \left(c_{m_2}^{(2)}, w_{m_2}^{(2)}\right)\right\} \end{aligned} \tag{2}$$

where m_1 and m_2 are the number of clusters in X_1 and X_2, respectively. We also define the ground distance matrix $G = [g_{ij}]$, where g_{ij} denotes the ground distance between $c_i^{(1)}$ and $c_j^{(2)}$. Any distance measure can be used for the ground distance.

We then find a flow matrix $F = [f_{ij}]$, where f_{ij} denotes the flow between the ith cluster of X_i and the jth cluster of X_2, that transforms X_1 into X_2 and minimizes the overall cost

$$Cost(X_1, X_2, F) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} g_{ij} f_{ij} \tag{3}$$

subject to

$$f_{ij} \geq 0 \quad 1 \leq i \leq m_i, \quad 1 \leq j \leq m_2 \tag{4}$$

$$\sum_{j=1}^{m_2} f_{ij} \leq w_i^{(1)} \quad 1 \leq i \leq m_1 \tag{5}$$

$$\sum_{i=1}^{m_1} f_{ij} \leq w_j^{(2)} \quad 1 \leq j \leq m_2 \tag{6}$$

$$\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} f_{ij} = \min\left(\sum_{i=1}^{m_1} w_i^{(1)}, \sum_{j=1}^{m_2} w_j^{(2)}\right). \tag{7}$$

The above optimization problem is a generic linear programming problem and can be efficiently solved [12]. Once we obtain the optimal flow matrix F, we can calculate the EMD as

$$EMD(X_1, X_2) \equiv \frac{\Sigma_{i=1}^{m_1} \Sigma_{j=1}^{m_2} g_{ij} f_{ij}}{\Sigma_{i=1}^{m_1} \Sigma_{j=1}^{m_2} f_{ij}}. \tag{8}$$

For texture classification and object recognition tasks, *normalized signatures* have commonly been used, i.e., the sum of all weights of a signature is 1 and all signatures have equal number of clusters. In this case, the EMD can be simply calculated as follows:

$$EMD(X_1, X_2) \equiv \sum_{i=1}^{m} \sum_{j=1}^{m} g_{ij} f_{ij}. \tag{9}$$

In this brief, we follow this tradition for a simpler formulation.

LDA utilizes label information to maximize the interclass distances and minimize the intraclass distances. More formally, LDA solves the following optimization problem to obtain the *optimal transformation matrix* $\hat{T}$ [5]:

$$\hat{T} = \arg\max_{T \in R^{m \times r}} \frac{\det(T^T S_B T)}{\det(T^T S_W T)} \tag{10}$$

where m and r are the dimensionality of feature space and embedding space, respectively, and S_W and S_B are the *within-classes scatter matrix* and the *between-classes scatter matrix*, respectively, which are defined as

$$S_W \equiv \sum_c \sum_{i \in c} n_c (x_i - \mu_c)(x_i - \mu_c)^T \tag{11}$$

$$S_B \equiv \sum_c n_c (\mu_c - \bar{x})(\mu_c - \bar{x})^T \tag{12}$$

where n_c is the number of samples in class c and

$$\mu_c = \frac{\Sigma_{i \in c} x_i}{n_c} \quad \text{and} \quad \bar{x} = \frac{\Sigma_c n_c \mu_c}{\Sigma_c n_c}. \tag{13}$$

Equation (10) can be cast into a generalized eigenvalue problem, whose solution $\hat{T}$ consists of generalized eigenvectors [21].

According to [8, eqs. (11) and (12)] can be reformulated in a pairwise manner as follows:

$$S_W = \frac{1}{2} \sum_c \sum_{i,j \in c} \frac{1}{n_c} (x_i - x_j)(x_i - x_j)^T \tag{14}$$

$$S_B = \frac{1}{2} \sum_{i,j} \frac{1}{n} (x_i - x_j)(x_i - x_j)^T - S_W \tag{15}$$

where n is the number of all samples.

Based on these pairwise formulations of LDA, LFDA [8] weights sample pairs of the same class in (14) and (15) with affinity between the pairs. Thanks to the affinity factors, LFDA can overcome the rank deficiency problem of classical LDA, and also can be applied to multimodal labeled data [8].

III. Signature-LDA

A. Formulation

In case of signatures, the center in a class (μ_c) and the center of the entire samples ($\bar{x}$) in (13) are not available because signatures are not on the Euclidean space. On the other hand, pairwise distances of signatures can be calculated using the EMD. Based on this idea, we reformulate (14) and (15) to be applied to signatures in this section.

The pairwise formulation in (14) and (15) can be generalized as

$$S_W = \frac{1}{2}\sum_c \sum_{i,j\in c} \frac{1}{n_c} P_{ij} \tag{16}$$

$$S_B = \frac{1}{2}\sum_{i,j} \frac{1}{n} P_{ij} - S_W \tag{17}$$

where P_{ij} is the *squared difference matrix* between the ith and the jth samples. When each sample is a vector, $P_{ij} = (x_i - x_j)(x_i - x_j)^T$ as in (14) and (15).

To devise LDA for signatures, we need P_{ij} that is suitable for signatures. By extending the manner of the original P_{ij}, we define P_{ij} for signatures as the weighted sum of the squared difference matrix between all pairs of (a center in X_i, a center in X_j), where the weights are the optimal flows of the EMD between X_i and X_j. A more formal definition is as follows.

Let the signature of the ith sample be $X_i = \{(c_1^{(i)}, w_1^{(i)}), (c_2^{(i)}, w_2^{(i)}), \ldots, (c_m^{(i)}, w_m^{(i)})\}$ and $F^{(ij)} = [f_{kl}^{(ij)}]$ be the optimal flow matrix of the EMD between two signatures X_i and X_j. We define the squared difference matrix of signature-LDA between the ith and the jth samples as

$$P_{ij} \equiv \sum_{k,l=1}^{m} f_{kl}^{(ij)} \left(c_k^{(i)} - c_l^{(j)}\right)\left(c_k^{(i)} - c_l^{(j)}\right)^T. \tag{18}$$

Equation (18) can be simplified in the matrix form as follows:

$$\begin{aligned}
P_{ij} &= \sum_{k,l=1}^{m} f_{kl}^{(ij)}\left(c_k^{(i)}c_k^{(i)T} - c_k^{(i)}c_l^{(j)T} - c_l^{(j)}c_k^{(i)T} + c_l^{(j)}c_l^{(j)T}\right)\\
&= \sum_{k,l=1}^{m} f_{kl}^{(ij)} c_k^{(i)}c_k^{(i)T} - \sum_{k,l=1}^{m} f_{kl}^{(ij)} c_k^{(i)}c_l^{(j)T}\\
&\quad - \sum_{k,l=1}^{m} f_{kl}^{(ij)} c_l^{(j)}c_k^{(i)T} + \sum_{k,l=1}^{m} f_{kl}^{(ij)} c_l^{(j)}c_l^{(j)T}\\
&= C^{(i)}D^{(ij)}C^{(i)T} - C^{(i)}F^{(ij)}C^{(j)T}\\
&\quad - C^{(j)}F^{(ij)T}C^{(i)T} + C^{(j)}E^{(ij)}C^{(j)T}
\end{aligned} \tag{19}$$

where $C^{(i)} \equiv [c_1^{(i)}|c_2^{(i)}|\cdots|c_m^{(i)}]$, and $D^{(ij)}$ and $E^{(ij)}$ are diagonal matrices and the kth diagonal element of $D^{(ij)}$ and the lth diagonal element of $E^{(ij)}$ are respectively shown as follows:

$$D_{kk}^{(ij)} \equiv \sum_{l=1}^{m} f_{kl}^{(ij)} \quad \text{and} \quad E_{ll}^{(ij)} \equiv \sum_{k=1}^{m} f_{kl}^{(ij)}. \tag{20}$$

Since $F^{(ij)^T} = F^{(ji)}$, we can easily show that $E^{(ij)} = D^{(ji)}$ from (20). Putting this result into (16) and (17) gives the within-class scatter matrix of signature-LDA as follows:

$$\begin{aligned}
S_W &= \frac{1}{2}\sum_c \sum_{i,j\in c} \frac{1}{n_c}\Big(C^{(i)}D^{(ij)}C^{(i)T} - C^{(i)}F^{(ij)}C^{(j)T}\\
&\qquad - C^{(j)}F^{(ij)T}C^{(i)T} + C^{(j)}E^{(ij)}C^{(j)T}\Big)\\
&= \frac{1}{2}\sum_c \sum_{i,j\in c} \frac{1}{n_c}\Big(C^{(i)}D^{(ij)}C^{(i)T} - C^{(i)}F^{(ij)}C^{(j)T}\\
&\qquad - C^{(j)}F^{(ji)}C^{(i)T} + C^{(j)}D^{(ji)}C^{(j)T}\Big)\\
&= \frac{1}{2}\sum_c \sum_{i,j\in c} \frac{1}{n_c}\Big(C^{(i)}D^{(ij)}C^{(i)T} - C^{(i)}F^{(ij)}C^{(j)T}\Big)\\
&\quad + \frac{1}{2}\sum_c \sum_{j,i\in c} \frac{1}{n_c}\Big(C^{(j)}D^{(ji)}C^{(j)T} - C^{(j)}F^{(ji)}C^{(i)T}\Big)\\
&= \sum_c \sum_{i,j\in c} \frac{1}{n_c}\Big(C^{(i)}D^{(ij)}C^{(i)T} - C^{(i)}F^{(ij)}C^{(j)T}\Big).
\end{aligned} \tag{21}$$

In the same way, the between-class scatter matrix of signature-LDA can be computed as

$$S_B = \sum_{i,j} \frac{1}{n}\Big(C^{(i)}D^{(ij)}C^{(i)T} - C^{(i)}F^{(ij)}C^{(j)T}\Big) - S_W. \tag{22}$$

Solving the generalized eigenvalue problem with above S_W and S_B yields eigenvalue/eigenvector pairs $(\lambda_1, \nu_1), \ldots, (\lambda_m, \nu_m)$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. In order to construct the optimal transformation matrix of signature-LDA, we adopt the weighting scheme of LFDA [8]. More specifically, the optimal transformation matrix is constructed with the eigenvectors weighted by the square root of the corresponding eigenvalues after normalizing the L2 norm of each eigenvector to 1 as follows:

$$\hat{T} \leftarrow \left[\sqrt{\lambda_1}\nu_1|\sqrt{\lambda_2}\nu_2|\cdots|\sqrt{\lambda_m}\nu_m\right]. \tag{23}$$

This scheme strengthens the influences of the major eigenvectors while weakening those of the minor ones so as to empirically show good performance.

B. Discussion on Signature-LDA

Signature-LDA can be interpreted as a kernel LDA [22] with the following kernel function:

$$K(X_i, X_j) = -\frac{1}{2}\sum_{k,l=1}^{m} f_{kl}^{(ij)}\left(c_k^{(i)} - c_l^{(j)}\right)^T\left(c_k^{(i)} - c_l^{(j)}\right). \tag{24}$$

Note that this kernel function is defined on a pair of signatures. This kernel trick maps signatures into a lower dimensional vector space whose dimension is the same as that of a cluster center. This is opposite to the general kernel trick that maps samples into a higher dimensional vector space. As a result, signature-LDA explicitly yields the optimal transformation matrix that can be directly applied to the center matrices of original signatures, as we will show in the next section. Such a transformation matrix cannot be obtained in general kernel LDA because neither the dimension of the kernel space conforms to that of the original feature space nor the nonlinear mapping is explicitly computed. In this respect, signature-LDA is distinct from general kernel LDA methods.

Signature-LDA can also be interpreted as a pairwise formulation of LDA with weights, when each center from a signature is considered as a sample. Mathematically, this is obtained by inserting (18) into (16) and (17) as follows:

$$S_W = \frac{1}{2}\sum_{c}\sum_{i,j\in c}\sum_{k,l=1}^{m}\frac{f_{kl}^{(ij)}}{n_c}\left(c_k^{(i)} - c_l^{(j)}\right)\left(c_k^{(i)} - c_l^{(j)}\right)^T \quad (25)$$

$$S_B = \frac{1}{2}\sum_{i,j}\sum_{k,l=1}^{m}\frac{f_{kl}^{(ij)}}{n}\left(c_k^{(i)} - c_l^{(j)}\right)\left(c_k^{(i)} - c_l^{(j)}\right)^T - S_W. \quad (26)$$

Comparing these equations with (14) and (15) reveals that (25) and (26) are a pairwise LDA formulation with weight $f_{kl}^{(ij)}$ on the squared difference matrix between the kth center of the ith signature and the lth center of the jth signature. Thus, for every pair of centers between which there is a positive flow, signature-LDA reduces the distance between the two centers when their labels are the same, or increases the distance when their labels are different.

The formulation of signature-LDA is similar to that of LFDA in that both are represented by the sum of weighted squared difference matrices. Apart from the fact that LFDA cannot be applied to signatures, the critical difference is in the weight factor, LFDA uses heuristically chosen affinity coefficients, but signature-LDA uses the optimal flows of the EMD, which represent the optimal proximity between cluster centers that minimizes the cost to transform a signature to the other. These carefully chosen weights enable signature-LDA to select more descriptive features.

Due to the similarity in its formulation, signature-LDA inherits the key strengths of LFDA. The most important is that signature-LDA can resolve the rank deficiency problem of classical LDA. When the number of classes in a given dataset is r, at most $r-1$ generalized eigenvectors are meaningful because the between-class scatter matrix S_B has at most rank $r-1$ [5]. On the other hand, since S_B in signature-LDA generally has full rank, the dimensionality of the embedding space is not limited by the number of classes in signature-LDA. In addition, signature-LDA can be extended to nonlinear discriminant analysis by adopting the kernel trick as LFDA does [8].

Signature-LDA also has advantages in computational implementation. Since the optimal flow is likely to be zero when the centers are farther apart, the optimal flow matrix F is generally sparse. This eases the computation of the weighted sum of the squared difference matrices. More specifically, since F generally contains $O(m)$ nonzero elements, S_W and S_B can be computed in $O(n^2(mp+p^2))$ rather than $O(n^2(m^2p+mp^2))$ where p is the size of a center vector. Since the generalized eigenvalue problem can be solved in $O(p^3)$, the overall time complexity of signature-LDA is $O(n^2(mp+p^2)+p^3)$. As signature-LDA has a closed-form matrix formulation, the computational efficiency can be boosted by specialized matrix computation tools. The optimality, the full rank property, and the computational sparsity give signature-LDA competitive advantages in handling signatures.

Similar to the EMD kernel [20], the kernel of signature-LDA has always constructed positive definite Gram matrices in our experiments although we do not have a proof of positive definiteness for the kernel.

C. EMD Calculation with Signature-LDA

Although signature-LDA can be used as a general feature selection method of local feature vectors, the relevance with the EMD generally leads signature-LDA to be used for more discriminative EMD calculation. To do this, the optimal transformation matrix $\hat{T}$ obtained through signature-LDA needs to be applied to project every center of signatures to the embedding space as follows:

$$C^{(i)} \leftarrow \hat{T}^T \times C^{(i)} \quad \text{for } i = 1,\ldots,n \quad (27)$$

where $C^{(i)}$ is the cluster center matrix of the signature generated from the ith image, each column of which is a cluster center.

We can omit the recalculation of the optimal flow matrices after signature-LDA by reusing the optimal flow matrices obtained during signature-LDA. This simplification allows the running time of EMD calculation with signature-LDA to be comparable to that of without signature-LDA. The pseudocode of the EMD calculation process with signature-LDA is shown in Algorithm 1.

IV. Experiments

In this section, we empirically show the performance of signature-LDA on three texture databases: Brodatz [23], UIUCTex [18], KTH-TIPS [24], and USC-SIPI Textures [25]. Through experiments, we show that signature-LDA not only improves signature/EMD-based texture classification performance, but also outperforms the other feature selection methods devised for local image features.

A. Experimental Setup

Brodatz [23] texture album contains 112 different texture classes each of which is represented by one image divided into nine subimages. UIUCTex [18] contains 1000 grayscale texture photographic images which are categorized into 25 classes with 40 samples for each. KTH-TIPS [24] contains 810 images of 10 classes for each of which three different illumination directions, three different poses, and nine different scales are applied, thus producing 81 images per class. USC-SIPI Textures [25] originally contains 154 texture images. Among them, except rotated and mosaic images, 58 images are divided into nine sub-images.

To extract local image features from each image, we adopt the Harris detector [26] for detection of the region of interest

TABLE I

MISCLASSIFICATION RATES (%) WITH/WITHOUT SIGNATURE-LDA ON BRODATZ, UIUCTEX, KTH-TIPS, AND USC-SIPI

Brodatz				UIUCTex			
training size	without signature-LDA	with signature-LDA	performance gain (p-value)	training size	without signature-LDA	with signature-LDA	performance gain (p-value)
2	28.49 ± 1.92	19.15 ± 2.72	**9.34 ± 1.74** (<0.0001)	5	13.07 ± 2.03	7.89 ± 0.90	**5.19 ± 2.31** (<0.0001)
3	16.44 ± 1.23	11.85 ± 1.50	**4.59 ± 1.62** (<0.0001)	10	5.83 ± 0.94	3.43 ± 0.74	**2.40 ± 1.28** (0.0002)
4	12.76 ± 1.01	9.06 ± 0.76	**3.69 ± 1.26** (<0.0001)	15	4.18 ± 0.83	2.26 ± 0.41	**1.92 ± 0.77** (<0.0001)
5	10.11 ± 1.04	7.36 ± 0.83	**2.75 ± 1.12** (<0.0001)	20	3.46 ± 1.07	1.76 ± 0.50	**1.70 ± 1.03** (0.0004)

KTH-TIPS				USC-SIPI			
training size	without signature-LDA	with signature-LDA	performance gain (p-value)	training size	without signature-LDA	with signature-LDA	performance gain (p-value)
10	21.31 ± 1.61	19.38 ± 2.22	**1.93 ± 2.37** (0.0185)	2	37.81 ± 2.33	33.89 ± 1.68	**3.92 ± 1.63** (<0.0001)
20	15.74 ± 1.34	13.82 ± 1.44	**1.92 ± 0.99** (0.0001)	3	27.73 ± 1.28	19.77 ± 1.33	**7.96 ± 1.30** (<0.0001)
30	14.37 ± 1.48	12.84 ± 0.98	**1.53 ± 1.90** (0.0194)	4	23.55 ± 1.55	14.79 ± 1.72	**8.76 ± 1.98** (<0.0001)
40	13.66 ± 2.02	11.68 ± 1.00	**1.98 ± 1.32** (0.0007)	5	19.48 ± 1.19	12.46 ± 1.21	**7.03 ± 1.58** (<0.0001)

Training size indicates the number of training sample per class. p-values are computed through paired-sample t-tests.

Algorithm 1 EMD calculation using signature-LDA

Input n: the number of sample images,
$\{Z_i\}$: a set of local image feature matrices,
$\{y_i\}$: a set of ground truth class ids,
Output $[V_{ij}]$: EMD matrix, (i, j) element of which is the EMD between ith and jth samples.

for $i = 1$ to n **do**
  Generate the signature X_i from Z_i.
end for
for $i, j = 1$ to n **do**
  Calculate the ground distance matrix $G^{(ij)}$.
  Calculate the optimal flow matrix $F^{(ij)}$.
end for
Compute S_W and S_B through (20), (21), and (22) based on training samples.
$(\lambda_{1:m}, \nu_{1:m}) \leftarrow$ the solution of the generalized eigenvalue problem $S_B\nu = \lambda S_W \nu$. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$.
Normalize eigenvectors $\nu_1, \ldots,$ and ν_m.
$\hat{T} \leftarrow [\sqrt{\lambda_1}\nu_1 | \sqrt{\lambda_2}\nu_2 | \cdots | \sqrt{\lambda_m}\nu_m]$.
for $i = 1$ to n **do**
  $C^{(i)} \leftarrow \hat{T}^T \times C^{(i)}$
end for
for $i, j = 1$ to n **do**
  Recalculate ground distance matrix $G^{(ij)}$.
  Compute V_{ij} based on $G^{(ij)}$ and $F^{(ij)}$.
end for

and the scale-invariant feature transform (SIFT) descriptor [14] for descriptor computation. Affine adaptation process [27] is then applied to achieve affine invariance.

For evaluation of signature-LDA and other methods, we randomly select a certain number of images per class as the training set and use the remaining images as the test set. After repeating 10 trials of test, the average and standard deviation of 10 misclassification rates are reported.

All methods were implemented in MATLAB, except the EMD computation module and the libsvm package which were implemented in C. For running time evaluation, we utilized a computer with a 2.66-GHz Intel CPU and a 2-GB RAM.

B. Comparison Between with and without Signature-LDA

We adopt the approach of the signature/EMD-based classification method using local image features [18], [20] as our testbed. Under the setup, we compare the performance of two cases: the testbed without signature-LDA (as the baseline) and that with signature-LDA.

The procedure of the experiment is summarized as follows: After computing SIFT descriptors, signatures are generated using K-means clustering for each image. For K-means clustering, we determined K by cross validation among 10, 20, and 40, the previous work recommended setting K between 20 to 40 [20]. Afterward, the EMDs are calculated between all pairs of signatures with the Euclidean distance as the ground distance. A one-against-one multiclass SVM classification follows with the EMD kernel [18] for classification. The SVM penalty parameter is determined through cross validation, the parameter is validated with values 2^k with $k = 1, \ldots, 9$. Before the EMD calculation, we selectively apply or do not apply signature-LDA to the signatures and compare the accuracies of the two cases.

Table I shows that signature-LDA improves upon the classification performance of the baseline approach regardless of the image database and training size. In all of the cases, the performance gains due to signature-LDA are statistically meaningful as shown by the p-values computed through paired-sample t-tests. Table II demonstrates the running time of EMD computation between one image and the others given an image sequence. This result reveals that signature-LDA has a reasonable additional computational cost compared to EMD computation without signature-LDA.

C. Comparison with Other Feature Selection Methods

To the best of our knowledge, there exists no feature selection method utilizing label information for local image features themselves or their signature representation, in this respect, our signature-LDA is original. Therefore, we compare our signature-LDA with other feature selection methods by applying to visual word histograms [13], which is one of the vector representations of local image features.

TABLE II

RUNNING TIMES OF EMD COMPUTATION PER IMAGE WITH/WITHOUT SIGNATURE-LDA ON BRODATZ, UIUCTEX, KTH-TIPS, AND USC-SIPI

Databases	without signature-LDA	with signature-LDA	increased time rate	Databases	without signature-LDA	with signature-LDA	increased time rate
Brodatz	1.94	2.12	9.28%	UIUCTex	3.08	3.19	3.57%
KTH-TIPS	1.65	1.72	4.24%	USC-SIPI	1.87	1.99	6.42%

Per class, 5 and 10 samples are used as training data for Brodatz/USC-SIPI and UIUCTex/KTH-TIPS, respectively.

TABLE III

PERFORMANCE COMPARISON WITH OTHER DIMENSIONALITY REDUCTION METHODS IN TERMS OF THE MISCLASSIFICATION RATE (%)

Databases	training size	signature+ signature-LDA	histogram+ PCA	histogram+ ICA	histogram+ MFA	histogram+ LDA	histogram+ LFDA
Brodatz	2	**19.15 ± 2.72**	30.93 ± 1.58	25.75 ± 1.54	30.51 ± 1.93	37.05 ± 2.39	41.48 ± 3.66
	3	**11.85 ± 1.50**	17.99 ± 0.95	16.91 ± 0.98	19.16 ± 0.76	16.82 ± 1.08	16.80 ± 1.19
	4	**9.06 ± 0.76**	13.62 ± 0.90	13.59 ± 0.85	12.85 ± 0.90	12.59 ± 1.01	13.03 ± 1.16
	5	**7.36 ± 0.83**	11.42 ± 1.41	11.35 ± 1.14	11.69 ± 3.46	9.59 ± 0.83	10.38 ± 1.37
UIUCTex	5	**7.89 ± 0.90**	10.13 ± 0.91	8.61 ± 0.62	8.65 ± 1.00	**8.38 ± 1.21**	12.93 ± 2.43
	10	**3.43 ± 0.74**	4.35 ± 0.47	**3.96 ± 0.70**	**3.95 ± 0.78**	4.07 ± 0.75	5.56 ± 1.11
	15	**2.26 ± 0.41**	3.30 ± 0.71	**2.64 ± 0.74**	2.77 ± 0.75	2.98 ± 0.81	3.73 ± 0.89
	20	**1.76 ± 0.50**	2.40 ± 0.85	**1.86 ± 0.48**	**1.60 ± 0.38**	**2.02 ± 0.65**	3.02 ± 0.80
KTH-TIPS	10	**19.38 ± 2.22**	21.10 ± 1.73	**20.41 ± 1.20**	21.07 ± 1.57	**19.62 ± 2.95**	20.79 ± 2.56
	20	**13.82 ± 1.44**	15.44 ± 2.26	**14.16 ± 1.69**	15.28 ± 1.99	**13.57 ± 1.15**	14.87 ± 1.59
	30	**12.84 ± 0.98**	13.76 ± 2.00	**13.18 ± 1.65**	**12.76 ± 1.85**	**12.55 ± 1.20**	13.73 ± 1.85
	40	**11.68 ± 1.00**	12.02 ± 1.13	**11.35 ± 1.14**	**11.78 ± 1.47**	**11.51 ± 1.27**	12.24 ± 2.09
USC-SIPI	2	**33.89 ± 1.68**	**34.06 ± 1.51**	34.56 ± 1.61	35.39 ± 1.31	45.76 ± 2.84	46.26 ± 3.08
	3	**19.77 ± 1.33**	26.52 ± 1.32	25.95 ± 1.55	28.05 ± 1.81	33.53 ± 1.99	26.28 ± 1.24
	4	**14.79 ± 1.72**	22.72 ± 2.04	23.17 ± 1.69	23.79 ± 2.10	28.66 ± 1.90	22.66 ± 0.93
	5	**12.46 ± 1.21**	22.54 ± 2.20	20.09 ± 1.99	22.76 ± 3.11	26.29 ± 1.51	21.42 ± 1.77

For each case, the best performance and comparable ones based on the t-test at the significance level of 5% are shown in bold.

The procedure of the histogram-based image classification is as follows. A vocabulary is constructed from local feature vectors using K-means clustering. We determine the size of vocabulary K by cross validation among 100, 300, 500, and 1000. For each local feature vector, the nearest visual word is replaced with the vector. A histogram is then generated for each image as a fraction vector indicating the frequency of each vocabulary in the image. After applying a feature selection method to the histograms, a radial basis function kernel-based SVM classification is performed. SVM parameters are determined by cross validation, the penalty parameter C and the kernel parameter γ are validated with values 2^k with $k = -2, \ldots, 6$ and $k = 4, \ldots, 12$, respectively.

We compare signature-LDA with other popular subspace learning algorithms: principal component analysis (PCA) [28], independent component analysis (ICA) [29], marginal Fisher analysis (MFA) [30], LDA [21], and LFDA [8]. For PCA, ICA, and MFA, the dimensionality of embedding space is determined by cross validation among 10, 20, 30, 50, 100, 200, and 300.

Table III shows that signature-LDA outperforms the other methods. First of all, signature representation is superior to histogram representation in terms of expression power because intrinsic information of local features are sacrificed in the process of histogram construction due to binning problems. In addition, PCA and ICA, which are classical unsupervised dimension reduction methods, do not show comparable classification power because they do not utilize label information. The other supervised methods are not comparable to signature-LDA as well. MFA and LFDA require sufficient number of training samples per class to reliably construct effective intrinsic and penalty graphs. LDA suffers from the rank deficiency problem, the dimension of the embedding space is always less than the number of classes, which often degrades the performance when the number of classes is small.

In addition to performance, signature-LDA has an advantage over the other methods in parameter tuning. All the other methods require parameters that need to be determined either heuristically or empirically. PCA, ICA, and MFA need to set the dimensionality of the embedding space. MFA also requires the number of the nearest neighbors to construct intrinsic and penalty graphs. LFDA needs to determine the scale parameter. On the other hand, signature-LDA is free from parameter tuning, which allows fast model learning.

V. CONCLUSION AND FUTURE WORK

We have proposed signature-LDA, which is a supervised feature selection method devised for signatures constructed from the local features of images. Signature-LDA has several good properties including avoiding the rank deficiency problem of classical LDA and requiring no parameter tuning. We empirically show the discriminating power of signature-LDA on texture databases. Signature-LDA can be easily applied to a new EMD variant [31] that can be computed in linear time. We leave the experimental validation as future work. Furthermore, we plan to extend the idea of signature-LDA to scene classification or object recognition by improving signature-based approaches to additionally utilize context or spatial information.

REFERENCES

[1] P. N. Belhumeour, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[2] S. Dudoit, J. Fridlyand, and T. P. Speed, "Comparison of discrimination methods for the classification of tumors using gene expression data," *J. Am. Statist. Assoc.*, vol. 97, no. 457, pp. 77–87, Mar. 2002.

[3] D. L. Swets and J. J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 831–836, Aug. 1996.

[4] C. Bauckhage, and J. K. Tsotsos, "Seperable linear discriminant classification," in *DAGM*, vol. 3663. Berlin, Germany: Springer-Verlag, 2005, pp. 318–325.

[5] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. Boston, MA: Academic, 1990.

[6] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Rio de Janeiro, Brazil, Oct. 2007, pp. 1–7.

[7] T.-K. Kim, S.-F. Wong, B. Stenger, J. Kittler, and R. Cipolla, "Incremental linear discriminant analysis using sufficient spanning set approximations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, Jun. 2007, pp. 1–8.

[8] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, May 2007.

[9] Z. Zhao, L. Sun, S. Yu, H. Liu, and J. Ye, "Multiclass probabilistic kernel discriminant analysis," in *Proc. 21st Int. Joint Conf. Artif. Intell.*, Pasadena, CA, 2009, pp. 1363–1368.

[10] J. Ye, R. Janardan, and Q. Li, "2-D linear discriminant analysis," in *Proc. NIPS*, vol. 17. 2004, pp. 1569–1576.

[11] S. Ji and J. Ye, "Generalized linear discriminant analysis: A unified framework and efficient model selection," *IEEE Trans. Neural Netw.*, vol. 19, no. 10, pp. 1768–1782, Oct. 2008.

[12] M. Crosier and L. D. Griffin, "Texture classification with a dictionary of basic image features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, Jun. 2008, pp. 2502–2508.

[13] T. Leung and J. Malik, "Recognizing surfaces using 3-D textons," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 29–44, 2001.

[14] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[15] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 524–531.

[16] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Proc. 9th Eur. Conf. Comput. Vis.*, vol. 4. Graz, Austria, May 2006, pp. 490–503.

[17] P. Sabzmeydani and G. Mori, "Detecting pedestrians by learning shapelet features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, Jun. 2007, pp. 511–518.

[18] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.

[19] Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, Nov. 2000.

[20] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 213–238, Jun. 2007.

[21] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Hoboken, NJ: Wiley, 2001.

[22] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Mullers, "Fisher discriminant analysis with kernels," in *Proc. IEEE Signal Process. Soc. Workshop*, Madison, WI, Aug. 1999, pp. 41–48.

[23] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*. New York: Dover, 1966.

[24] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh, "On the significance of real-world conditions for material classification," in *Proc. Eur. Conf. Comput. Vis.*, vol. 4. Prague, Czech Republic, 2004, pp. 253–266.

[25] *Signal and Image Processing Institute Datebase: Volume 1: Textures.* Electr. Eng. Dept., Univ. Southern California [Online]. Available: http://sipi.usc.edu/database/database.cgi?volume=textures

[26] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vis. Conf.*, Manchester, U.K., 1988, pp. 147–151.

[27] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proc. 7th Eur. Conf. Comput. Vis.*, Copenhagen, Denmark, 2002, pp. 128–142.

[28] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag, 2002.

[29] A. Hyvrinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[30] D. Xu, S. Yan, D. Tao, S. Lin, and H. J. Zhang, "Marginal fisher analysis and its variants for human gait recognition and content-based image retrieva," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2811–2821, Nov. 2007.

[31] O. Pele and M. Werman, "A linear time histogram metric for improved SIFT matching," in *Proc. 10th Eur. Conf. Comput. Vis.*, Marseille, France, 2008, pp. 495–508.