

Ensemble Manifold Rank Preserving for Acceleration-Based Human Activity Recognition

Dapeng Tao, Lianwen Jin, *Member, IEEE*, Yuan Yuan, *Senior Member, IEEE*, and Yang Xue

Abstract—With the rapid development of mobile devices and pervasive computing technologies, acceleration-based human activity recognition, a difficult yet essential problem in mobile apps, has received intensive attention recently. Different acceleration signals for representing different activities or even a same activity have different attributes, which causes troubles in normalizing the signals. We thus cannot directly compare these signals with each other, because they are embedded in a nonmetric space. Therefore, we present a nonmetric scheme that retains discriminative and robust frequency domain information by developing a novel ensemble manifold rank preserving (EMRP) algorithm. EMRP simultaneously considers three aspects: 1) it encodes the local geometry using the ranking order information of intraclass samples distributed on local patches; 2) it keeps the discriminative information by maximizing the margin between samples of different classes; and 3) it finds the optimal linear combination of the alignment matrices to approximate the intrinsic manifold lied in the data. Experiments are conducted on the South China University of Technology naturalistic 3-D acceleration-based activity dataset and the naturalistic mobile-devices based human activity dataset to demonstrate the robustness and effectiveness of the new nonmetric scheme for acceleration-based human activity recognition.

Index Terms—Acceleration signals, frequency domain information, human activity recognition, intrinsic manifold approximation, rank order, spectral geometry.

I. INTRODUCTION

HUMAN activity recognition receives intensive attentions in recent years, due to many practical applications, such as video surveillance [39], [44], health care [34], [43], and context-aware computing [28], [47]. In general, pattern recognition schemes can directly handle the samples which are represented in a vector space. In most neural networks system,

Manuscript received September 25, 2013; revised September 3, 2014; accepted September 8, 2014. This work was supported in part by the National Science and Technology Support Plan under Grant 2013BAH65F01 and Grant 2013BAH65F04, in part by the Guangdong Natural Science Funds under Grant S2011020000541 and Grant S2012040008016, in part by the Shaanxi Key Innovation Team of Science and Technology under Grant 2012KCT-04, in part by the Guangdong Scientific and Technology Research Plan under Grant 2012A010701001, in part by the Research Fund for the Doctoral Program of Higher Education of China under Grant 20120172110023, and in part by the National Natural Science Foundation of China under Grant 61075021, Grant 61125106, Grant 61201348, Grant 61172143, and Grant 61472144.

D. Tao, L. Jin, and Y. Xue are with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: dapeng.tao@gmail.com; lianwen.jin@gmail.com; yxue@scut.edu.cn).

Y. Yuan is with the Center for OPTical IMagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China (e-mail: yuany@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNNLS.2014.2357794

such as traffic signs recognition [12] and character recognition [27], the samples can be easily converted into feature vectors after normalizing the size of the input images. However, existing human activity recognition systems, which can be classified into two groups: computer vision-based systems and accelerometer-based systems, often suffer from the difficulty of effective feature representation and selection. Computer vision-based human activity recognition systems [24] perform poorly because of the variations of illumination conditions in sophisticated environments and motion blur. Accelerometer-based human activity recognition systems [28] cannot perform well because the sensor positions are not fixed. We review some representative works as follows.

Computer vision-based human activity recognition systems [24] decomposed images into a variable number of parts using interest point/region detectors, and then some feature descriptors, such as scale invariant feature transform [32], speeded up robust features [3], and local binary patterns [33], are used to represent these parts. The conventional pattern recognition algorithms cannot apply. Spatial pyramid matching [29] and locality-constrained linear coding [46] are proposed to address this problem by utilizing efficient coding schemes. Using dissimilarity-based representation concepts, Carli *et al.* [10] proposed a common framework to obtain robust performance.

Accelerometer-based human activity recognition [28], [43], [47] is an important alternative, which has received increasing attentions due to the popularity of smart phones in recent years. It exploits acceleration signals from smart phones equipped with the accelerometer to analyze and recognize daily human activities such as standing, walking and running. This approach has largely increased the possibility to reach the dream of pervasive computing. Recently, Kwapisz *et al.* [28], Tentori and Favela [43], and Wang *et al.* [47] have made significant progress in mobile music recommendation and mobile activity monitor by utilizing a set of related technologies. Before we apply learning models to the acceleration signals, there is a problem to be noticed, i.e., different acceleration signals for representing different activities or even the same activity with different attributes that could cause trouble in aligning time-axis and determining the length of signals. However, it is a standard scheme [2], [28], [45] to use fixed-length signal segments to represent a human activity. It is difficult to directly compare the envelope of acceleration signals in testing stage, because they are embedded in a nonmetric space. Representative nonmetric learning algorithms are summarized as follows. Jacobs and Weinshall [25] proposed

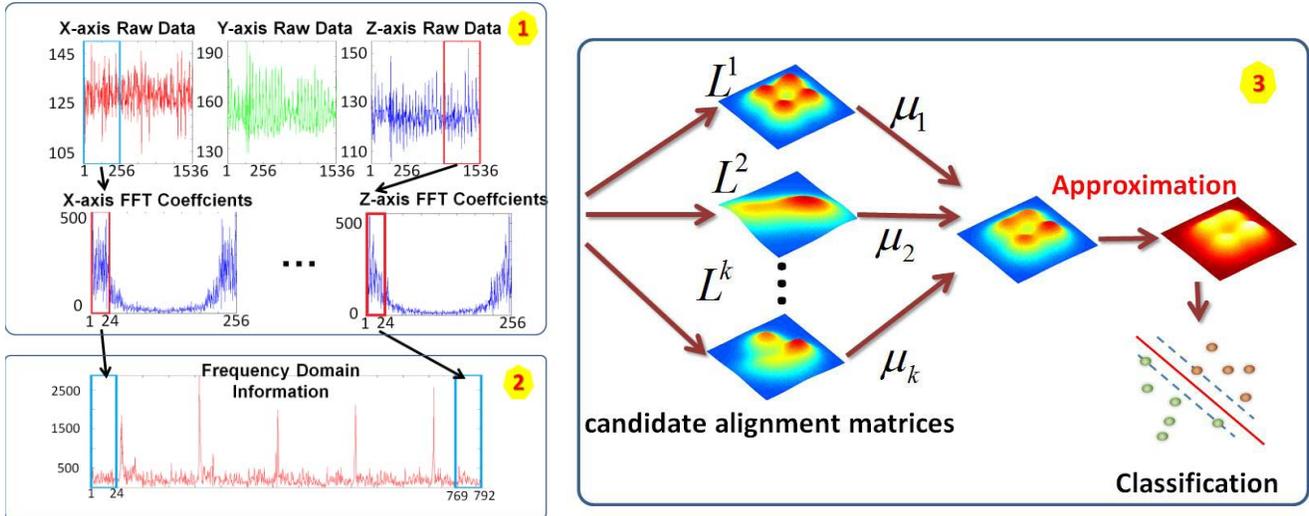


Fig. 1. Nonmetric scheme for acceleration-based human activity recognition. This scheme contains the following three components. 1—Using a fixed-length of window to cut out a section of the signal for extracting FFT coefficients and repeating this process from the beginning to the end of the signal. 2—Concatenating all the FFT coefficients as the frequency domain information. 3—Extracting the discriminative and robust information by EMRP for the subsequent classification.

a nonmetric distance-based classification scheme and applied it to image retrieval. Tosato *et al.* [44] proposed a weighted array of covariances to represent a tiny image. Xu *et al.* [49] introduced the Ricci flow to embed and rectify nonmetric dissimilarity data. Buló and Pelillo [7] proposed a model based on game theory to explore the hypergraph clustering problem.

Recently, various approaches based on accelerometer have been proposed to resolve the human activity recognition. Bonomi *et al.* [1] proposed a scheme that utilized a decision tree to classify measurements about time-domain information (including the average, the standard deviation, etc.) of a fixed intervals of signals. Ermers *et al.* [16] proposed a scheme, which combined a decision tree model with a neural networks model to improve the performance of acceleration-based human activity recognition. Recently, it has been found in experiments [50] that fast Fourier transform (FFT) coefficients obtain satisfactory recognition accuracy. Long *et al.* [31] utilized principal component analysis (PCA) [23] to find a low-dimensional representation for the time-domain information and the frequency domain information. However, PCA ignores the class labels that are essential to retain the discriminative information for the subsequent classification. Dimension reduction algorithms exploring the local geometric information find subspaces where the geodesic distance between the original signals can be effectively preserved. Therefore, we propose a new spectral geometry algorithm for human activity recognition, termed ensemble manifold rank preserving (EMRP). It models the similarity by exploiting the ranking order information of samples distributed on local patches and retains the discriminative information by maximizing the margin between samples of different classes. Especially, EMRP finds the optimal linear combination of the alignment matrices to approximate the intrinsic manifold lied in the data. Note that EMRP is different from other manifold learning based dimension reduction algorithms, such as rank preserving discriminant analysis (RPDA) [40], locality sensitive

discriminant analysis (LSDA) [9], and discriminative locality alignment (DLA) [52]. The main differences are twofold.

- 1) The performance of RPDA, LSDA, and DLA heavily rely on the predefined hyperparameters and cross validation is required for these parameters to obtain reasonable performance, because different hyperparameters induce different manifolds for encoding the local geometry. The proposed EMRP explores a set of hyperparameters and is robust to the initial hyperparameters by assuming that a linear combination of these manifolds approximates the intrinsic manifold.
- 2) RPDA, LSDA, and DLA use an alignment matrix to estimate the intrinsic manifold and can be solved by the generalized eigenvalue decomposition. However, EMRP use the optimal linear combination of the alignment matrices to approximate the intrinsic manifold. The solution of EMRP is an alternating minimization procedure which is a combination of the generalized eigenvalue decomposition and the coordinate descent.

For acceleration-based human activity recognition, we present a new nonmetric scheme that includes the following steps: 1) using a fixed length of window to cut out a section of the signal for extracting FFT coefficients and repeating this process from the beginning to the end of the signal; 2) concatenating all the FFT coefficients as the frequency domain information; and 3) extracting the discriminative and robust information by EMRP for the subsequent classification. Given the limited page length, we mainly describe the newly proposed EMRP, the other parts are easy to rebuild based on the references cited therein and will be sketched at the experimental part. Fig. 1 shows the architecture of the proposed scheme. The main contribution of this paper include: 1) we propose a spectral geometry approach EMRP which finds the optimal linear combination of the alignment matrices to avoid the time-consuming cross-validation procedure and 2) we conduct comprehensive experiments to demonstrate that

TABLE I
IMPORTANT NOTATIONS USED IN THIS PAPER AND THEIR DESCRIPTION

Notation	Description	Notation	Description
X	high-dimensional feature vectors	S_i	selection Matrix
Y	achieved succinct representations	$(w_i)_j$	weight factor of intra-class on a local patch
D	dimension of original feature vectors	$(v_i)_j$	weight factor of inter-class on a local patch
d	reduced dimension	γ	trade-off parameter in a local patch
N	size of the feature vectors X	L_i	representation of part optimization
k_1	number of closest intra-class samples	L	alignment matrix
k_2	number of closest inter-class samples	μ_k	weight factor of candidate alignment matrices
C_i	class label	m	number of candidate alignment matrices
X_i	local patch	β	trade-off parameter to control the contributions of the regularization item
Y_i	succinct representations of local patch	U	projection matrix

the newly developed EMRP can improve the performance of accelerometer-based human activity recognition.

The rest of this paper is organized as follows. In Section II, we detail the newly proposed EMRP. Section III shows the experimental results on two datasets. Section IV concludes this paper.

For convenience, Table I lists the important notations used in this paper.

II. RELATED WORK

In this section, we briefly review the existing techniques of spectral geometry. These techniques are widely used in accelerometer-based human activity recognition and other pattern recognition problems. We can simply group these techniques into two categories: globally linear learning based algorithms and manifold learning based algorithms.

The most representative works of globally linear learning include PCA [23] and linear discriminant analysis (LDA) [17]. The PCA aims to find the principal subspace in which the variance of the projected feature vectors is maximized. By utilizing class labels and assuming data from different classes are drawn from equal variance Gaussian distributions, the LDA finds the subspace in which the classes of objects can be properly separated. Ji and Ye [26] proposed a framework to unify LDA and some modified versions. Geometric mean-based subspace selection (GMSS) [41] have explored the general mean concept instead of the arithmetic mean used in LDA and achieved a superior performance.

Recently, manifold learning based algorithms play an important role in many intelligent systems and have received extensive attentions. In general, most manifold learning algorithms assume that the samples lie on a low dimensional submanifold embedded in a high-dimensional Euclidean space. The early representative works include ISOMAP [42], locally linear embedding [38], Laplacian eigenmaps (LEs) [4], and locality preserving projections [22]. Linear and supervised learning algorithms, such as LSDA [9], marginal Fisher's analysis (MFA) [48], and DLA [52] are popular for

classification problems. Recent research work explores the nonnegative and the sparse properties, e.g., nonnegative local coordinate factorization [11], graph regularized nonnegative matrix factorization [8], and nonnegative patch alignment framework (PAF) [19], [20].

III. ENSEMBLE MANIFOLD RANK PRESERVING

In this section, we present a new spectral geometry algorithm for human activity categorization, EMRP.

It is well known that the concentration of measure phenomenon significantly impacts on the outcomes of spectral geometry algorithms [14], [37], because the dissimilarities between pairs of intraclass samples are commonly indistinguishable, after dimension reduction based on spectral geometry algorithms [35]. Some unsupervised learning methods [36], [51], [53], which utilize preservation of the rank order information of samples [30] in the process of transforming samples from a high-dimensional space to a low-dimensional subspace, have been proposed to reduce this problem.

Supervised learning methods which utilize the class label information are suitable for classification tasks. They confront the different problem that dimension reduction always brings variations in the original distribution of samples [14]. Thus, ignoring the information about the interclass samples ranking order is preferred in the process of extraction of discriminant information [9], [52]. This arrangement is more practicable for us to selectively shrink or stretch a suitable distribution for original data.

Note that many spectral geometry approaches [9], [21] can be built under the umbrella of the PAF [52], because the intrinsic differences mainly exist in the part optimization stage and an almost identical process is provided in the whole alignment stage. We can conveniently use this characteristic to encode the rank information of intraclass samples and the discriminative information of interclass samples on local patches.

In human activity categorization, N human activities represented by a fixed length of features are recorded in a matrix

$X = [x_1, x_2, \dots, x_N] \in R^{D \times N}$, in which each D -dimensional feature vector $x_i \in R^D$ is associated with a class label (an entry in the label vector) in $C = l[c_1, c_2, \dots, c_N] \in Z^N$. The supervised spectral geometry algorithm aims to find a matrix $U \in R^{D \times d}$ that projects the original samples X from the high-dimensional space R^D to a low-dimensional subspace R^d and achieves succinct representations $Y = U^T X = [y_1, y_2, \dots, y_N] \in R^{d \times N}$, where $d < D$.

Under PAF [52], a patch defines a local coordinate system by exploring a sample and its nearest neighbors (NNs), and thus we denote a local patch $X_i = [x_i, x_{i1}, \dots, x_{i k_1}, x_{i1}, \dots, x_{i k_2}] \in R^{D \times (k_1 + k_2 + 1)}$ by utilizing an arbitrary sample x_i and its k_1 closest intraclass samples $x_{i1}, \dots, x_{i k_1}$ and k_2 closest interclass samples $x_{i1}, \dots, x_{i k_2}$. Thus, EMRP finds the corresponding low-dimensional representation, i.e., $Y_i = [y_i, y_{i1}, \dots, y_{i k_1}, y_{i1}, \dots, y_{i k_2}] \in R^{d \times (k_1 + k_2 + 1)}$. The optimization objectives of patches are: 1) to preserve the ranking order information of the intraclass samples as much as possible and 2) to maximize the discriminative information between the interclass samples.

A. Ranking Order Information Preservation

Rank order information contributes to preserving the distribution information of samples [30]. However, it is difficult to directly model the rank information, therefore the relative distances are considered in the objective function. Motivated by the success of LE [5], we define the ranking order information of intraclass on a local patch as

$$R(y_i) = \sum_{j=1}^{k_1} \|y_i - y_{ij}\|^2 (w_i)_j \quad (1)$$

where the weight factor $(w_i)_j$ is

$$(w_i)_j = \begin{cases} \exp(-\|x_i - x_{ij}\|^2/t), & \text{if } x_{ij} \in N_{k_1}(x_i) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

the penalized factor $(w_i)_j$ models the difference by respectively utilizing a large and a small value of weighting to emphasize small and large distances in the original high-dimensional space.

In the frame of PAF, we further deduce (1) to

$$R(y_i) = \text{tr} \left\{ \begin{bmatrix} (y_i - y_{i1})^T \\ \vdots \\ (y_i - y_{ij})^T \end{bmatrix} \text{diag}(w_i) [y_i - y_{i1}, \dots, y_i - y_{ij}] \right\} \\ = \text{tr}(Y_{R(i)} L_{R(i)} Y_{R(i)}^T) \quad (3)$$

where

$$L_{R(i)} = \begin{bmatrix} -e_{k_1}^T \\ I_{k_1} \end{bmatrix} \text{diag}(w_i) [-e_{k_1} \quad I_{k_1}], e_{k_1} = \begin{bmatrix} \overbrace{1, \dots, 1}^{k_1} \end{bmatrix}^T \\ I_{k_1} = \text{diag}(\overbrace{1, \dots, 1}^{k_1}), \text{ and } Y_{R(i)} = [y_i, y_{i1}, \dots, y_{i k_1}].$$

B. Discriminative Information Maximization

In supervised spectral geometry learning schemes, discriminative information plays a critical role. In our EMRP scheme, we consider to extract the discriminative information by maximizing the sum of distances between y_i and its k_2 interclass samples

$$D(y_i) = \sum_{p=1}^{k_2} \|y_i - y_{ip}\|^2 (v_i)_j. \quad (4)$$

It is unavoidable that the original distribution varies in the procedure of dimension reduction [13], so we consider to drop the interclass ranking informant and define the weight factor $(v_i)_j$ as

$$(v_i)_j = \begin{cases} 1, & \text{if } x_{ij} \in N_{k_1}(x_i) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

in the frame of PAF, we further deduce (4) to

$$D(y_i) = \text{tr} \left\{ \begin{bmatrix} (y_i - y_{i1})^T \\ \vdots \\ (y_i - y_{ip})^T \end{bmatrix} \text{diag}(v_i) [y_i - y_{i1}, \dots, y_i - y_{ip}] \right\} \\ = \text{tr}(Y_{D(i)} L_{D(i)} Y_{D(i)}^T) \quad (6)$$

where

$$L_{D(i)} = \begin{bmatrix} -e_{k_2}^T \\ I_{k_2} \end{bmatrix} \text{diag}(v_i) [-e_{k_2} \quad I_{k_2}], e_{k_2} = \begin{bmatrix} \overbrace{1, \dots, 1}^{k_2} \end{bmatrix}^T \\ I_{k_2} = \text{diag}(\overbrace{1, \dots, 1}^{k_2}), \text{ and } Y_{D(i)} = [y_i, y_{i1}, \dots, y_{i k_2}].$$

Thus, we can obtain the optimization function on a local patch

$$\arg \min_{y_i} \left(\sum_{j=1}^{k_1} \|y_i - y_{ij}\|^2 (w_i)_j - \gamma \sum_{p=1}^{k_2} \|y_i - y_{ip}\|^2 (v_i)_j \right) \quad (7)$$

where $\gamma \in [0, 1]$ is a tradeoff parameter to integrate the contributions of intraclass samples and those of the interclass samples in a local patch. We denote the index set $F_i = \{i, i^1, i^2, \dots, i^{k_1}, i_1, i_2, \dots, i_{k_2}\}$ and deduce (7) to

$$\arg \min_{y_i} \sum_{j=1}^{k_1} \|y_i - y_{ij}\|^2 (w_i)_j - \gamma \sum_{p=1}^{k_2} \|y_i - y_{ip}\|^2 (v_i)_j \\ = \arg \min_{y_i} \sum_{j=1}^{k_1 + k_2} \|y_{F_i\{1\}} - y_{F_i\{j+1\}}\|^2 (w'_i)_j \\ = \arg \min_{Y_i} \text{tr} \left\{ Y_i \begin{bmatrix} -e_{k_1 + k_2}^T \\ I_{k_1 + k_2} \end{bmatrix} \text{diag}(w'_i) [-e_{k_1 + k_2} \quad I_{k_1 + k_2}] Y_i^T \right\} \quad (8)$$

where $\text{tr}(\cdot)$ is the trace operator

$$e_{k_1+k_2} = [1, \dots, 1]^T \in R^{k_1+k_2}, \quad I_{k_1+k_2} = \text{diag} \left(\overbrace{1, \dots, 1}^{k_1+k_2} \right)$$

$$w'_i = \left[\overbrace{(w_i)_1, \dots, (w_i)_{k_1}}^{k_1}, \overbrace{-\gamma_{k_1+1}, \dots, -\gamma_{k_2}}^{k_2} \right]$$

$$Y_i = [y_i, y_{i1}, \dots, y_{i k_1}, y_{i1}, \dots, y_{i k_2}].$$

In the frame of PAF, we further define the alignment matrix

$$L_i = \begin{bmatrix} -e_{k_1+k_2}^T \\ I_{k_1+k_2} \end{bmatrix} \text{diag}(w'_i) \begin{bmatrix} -e_{k_1+k_2} & I_{k_1+k_2} \end{bmatrix} \quad (9)$$

and (8) can be rewritten as

$$\arg \min_{Y_i} \text{tr}(Y_i L_i Y_i^T). \quad (10)$$

We denote the selection matrix as

$$(S_i)_{pq} = \begin{cases} 1, & \text{if } p = F_i\{q\} \\ 0, & \text{else} \end{cases} \quad (11)$$

where $S_i \in R^{N \times (k_1+k_2+1)}$. Afterward, we can integrate all local patches $Y_i = [y_i, y_{i1}, \dots, y_{i k_1}, y_{i1}, \dots, y_{i k_2}]$ in a unified coordinate system as follows. First, the coordinate of Y_i is selected from the global coordinate system. The coordinate of the low dimensional representation Y_i is then given by $Y = U^T X = [y_1, y_2, \dots, y_N] \in R^{d \times N}$

$$Y_i = Y S_i. \quad (12)$$

Thus, by utilizing (12), the optimization function on a local patch (10) can be rewritten as

$$\arg \min_Y \text{tr}(Y S_i L_i S_i^T Y^T). \quad (13)$$

Second, we sum over all the optimization functions on the local patches defined in (13) over N samples to obtain the whole alignment objective function and then have

$$\arg \min_Y \sum_{i=1}^N \text{tr}(Y S_i L_i S_i^T Y^T) = \arg \min_Y \text{tr}(Y L Y^T) \quad (14)$$

where $L = \sum_{i=1}^N S_i L_i S_i^T \in R^{N \times N}$ is the alignment matrix which: 1) preserves the ranking order information of the intra-class samples and 2) maximizes the discriminative information between the interclass samples.

However, different hyperparameters [e.g., k_1 in (1) and k_2 in (4)] induce different manifolds for encoding different geometric information, and thus seriously impact the performance metrics, e.g., the classification accuracy [5], [6]. In our EMRP scheme, the hyperparameters t , k_1 and k_2 are no trivial to choose the optimal settings which will affect the objective function (11). Thus, an automatic manifold approximation method is valuable to tackle this problem.

C. Ensemble Manifold for Hyperparameters Approximation

It is not easy to directly find the optimal settings of hyperparameters according to (14). Motivated by the assumption [18] that the intrinsic manifold can be approximated by a set of candidate manifolds, we can redefine the alignment matrix

$$L = \sum_{k=1}^m \mu_k L^k$$

$$\text{s.t. } \sum_{k=1}^m \mu_k = 1, \mu_k \geq 0, \quad k = 1, \dots, l \quad (15)$$

where the alignment matrix L^k is obtained by setting particular values of hyperparameters. Therefore, we can transform tuning hyperparameters of L in (14) into the problem of finding the optimal linear combination of candidate manifolds in (15).

The objective function incorporated with the automatic manifold approximation can then be written as

$$\arg \min_{Y, \mu} \text{tr} \left(Y \left(\sum_{k=1}^m \mu_k L^k \right) Y^T \right) + \beta \| \mu \|^2$$

$$\text{s.t. } \sum_{k=1}^m \mu_k = 1, \mu_k \geq 0, \quad k = 1, \dots, l \quad (16)$$

where we introduce the regularization term $\| \mu \|^2$ to smooth the weights μ_k to avoid select only one manifold, and β is the tradeoff parameter to control the contributions of the regularization item.

We can utilize the alternating optimization method to obtain a locally optimal solution of (16).

In particular, given a fixed μ , (16) reduces to

$$\arg \min_Y \text{tr} \left(Y \left(\sum_{k=1}^m \mu_k L^k \right) Y^T \right). \quad (17)$$

It is straightforward to show that (14) is equivalent to

$$\arg \min_U \text{tr} \left(U^T X \left(\sum_{k=1}^m \mu_k L^k \right) X^T U \right). \quad (18)$$

We introduce the constraint $U U^T = I_d$. The solution can be obtained by means of the generalized eigenvector problem

$$X \left(\sum_{k=1}^m \mu_k L^k \right) X^T u = \lambda u. \quad (19)$$

For a fixed U , the problem is reduced to

$$\arg \min_{\mu \in R^m} \text{tr} \left(Y \left(\sum_{k=1}^m \mu_k L^k \right) Y^T \right) + \beta \| \mu \|^2$$

$$\text{s.t. } \sum_{k=1}^m \mu_k = 1, \quad \mu_k \geq 0. \quad (20)$$

We further deduce (20) to

$$\arg \min_{\mu \in R^m} \sum_{k=1}^m \mu_k \text{tr} \left(Y \left(\sum_{k=1}^m L^k \right) Y^T \right) + \beta \sum_{k=1}^m \mu_k^2$$

$$\text{s.t. } \sum_{k=1}^m \mu_k = 1, \quad \mu_k \geq 0. \quad (21)$$

The coordinate descent algorithm can be adopted to optimize μ . With respect to the constraints $\sum_{k=1}^m \mu_k = 1, \mu_k \geq 0$, in each iteration, two elements are selected for updating while the others are fixed. Suppose μ_i and μ_j are prepared to be updated and the others are fixed. In terms of the constraint, we have

$$\mu_j = 1 - \sum_{\substack{k=1 \\ k \neq i, j}}^m \mu_k - \mu_i. \quad (22)$$

Let $T(\mu_i)$ denote the objective function

$$\begin{aligned} T(\mu_i) &= \sum_{\substack{k=1 \\ k \neq i, j}}^m \mu_k \text{tr}(YL^k Y^T) + \beta \sum_{\substack{k=1 \\ k \neq i, j}}^m \mu_k^2 + \mu_i \text{tr}(YL^i Y^T) \\ &\quad + \mu_j \text{tr}(YL^j Y^T) + \beta(\mu_i^2 + \mu_j^2) \\ &= \sum_{\substack{k=1 \\ k \neq i, j}}^m \mu_k \text{tr}(YL^k Y^T) + \beta \sum_{\substack{k=1 \\ k \neq i, j}}^m \mu_k^2 + \mu_i \text{tr}(YL^i Y^T) \\ &\quad + \left(1 - \sum_{\substack{k=1 \\ k \neq i, j}}^m \mu_k - \mu_i \right) \text{tr}(YL^j Y^T) \\ &\quad + \beta \left[\mu_i^2 + \left(1 - \sum_{\substack{k=1 \\ k \neq i, j}}^m \mu_k - \mu_i \right)^2 \right]. \end{aligned} \quad (23)$$

Take the derivative of T with respect to μ_i , we have

$$\begin{aligned} \frac{\partial T(\mu_i)}{\partial \mu_i} &= \text{tr}(YL^i Y^T) + (-1)\text{tr}(YL^j Y^T) \\ &\quad + \beta \left(2\mu_i + 2 \left(1 - \sum_{\substack{k=1 \\ k \neq i, j}}^m \mu_k - \mu_i \right) \cdot (-1) \right) \\ &= \text{tr}(YL^i Y^T) - \text{tr}(YL^j Y^T) + 2\beta(\mu_i - \mu_j) = 0. \end{aligned} \quad (24)$$

According to (24), we can obtain

$$\mu_i^* - \mu_j^* = \frac{1}{2\beta} \left(\text{tr}(YL^j Y^T) - \text{tr}(YL^i Y^T) \right). \quad (25)$$

By considering $\mu_i^* + \mu_j^* = \mu_i + \mu_j$, the updated μ_i^* can be expressed as

$$\mu_i^* = \frac{1}{4\beta} \left(\text{tr}(YL^j Y^T) - \text{tr}(YL^i Y^T) \right) + \frac{\mu_i + \mu_j}{2}. \quad (26)$$

With respect to the constraint $\mu_i \geq 0$, The solution of (20) can be divided into the following situations.

Algorithm 1 Ensemble Manifold Rank Preserving

	Training set $X = [x_1, x_2, \dots, x_N] \in R^{D \times N}$;
Input:	Class label $C_i \in Z^n$ d : dimension of the reduced space.
Output:	Orthogonal projection matrix $U = [u_1, u_2, \dots, u_d] \in R^{D \times d}$
Step 1:	Compute the whole alignment matrix according to $L^k = \sum_{i=1}^N S_i L_i S_i^T \in R^{N \times N}$;
Step 2:	Given a fixed μ , we can compute the projection matrix U by using (19).
Step 3:	For a fixed U , we optimize μ by using (27)–(29).

- 1) If $1/4\beta(\text{tr}(YL^j Y^T) - \text{tr}(YL^i Y^T)) + \mu_i + \mu_j/2 \leq 0$, we have

$$\begin{cases} \mu_i^* = 0 \\ \mu_j^* = \mu_i + \mu_j. \end{cases} \quad (27)$$

- 2) With respect to the symmetry of i and j , if $(1/4\beta)(\text{tr}(YL^i Y^T) - \text{tr}(YL^j Y^T)) + \mu_i + \mu_j/2 \leq 0$, we have

$$\begin{cases} \mu_i^* = \mu_i + \mu_j \\ \mu_j^* = 0. \end{cases} \quad (28)$$

- 3) Otherwise

$$\begin{cases} \mu_i^* = \frac{1}{4\beta} (\text{tr}(YL^j Y^T) - \text{tr}(YL^i Y^T)) + \frac{\mu_i + \mu_j}{2} \\ \mu_j^* = \mu_i + \mu_j - \mu_i^*. \end{cases} \quad (29)$$

We can use (27)–(29) to optimize the pairs of elements in μ until the objective function (20) does not decrease.

Based on the above discussions, we summarize EMRP in Algorithm 1.

IV. EXPERIMENTAL RESULTS

In this section, we first conducted the experiments of human behavior recognition on the widely used South China University of Technology (SCUT) naturalistic 3-D acceleration-based activity (SCUT NAA) datasets [50]. Considering the lack of standard human behavior recognition dataset based on mobile devices, we collected some human behavior activity samples to create a dataset by utilizing smart phones equipped with a 3-D acceleration sensor, namely, naturalistic mobile-devices based human activity (NMHA) datasets. For each sample, the FFT coefficients are extracted to represent the raw acceleration data. We evaluate the performance of the proposed method using the average recognition rates for each human activity. We also reported the confusion matrix for better understanding when the method fails. Further details of the experimental setup and baseline methods are given below.

A. Datasets

The widely used SCUT NAA dataset is publically available: <http://www.hcii-lab.net/data/>. It consists of 1278 samples, obtained from 44 people (34 males and 10 females). A tri-axial accelerometer located on a fixed position is used to gather 10 categories of human activity. Table II [50] lists all the

TABLE II
DEFINITION OF THE TEN KINDS OF ACTIVITIES [50]

Classes	Definition
Relaxing	Sitting & doing nothing
Walking	Walking 50 m at normal speed
Walking quickly	Walking 50 m faster than normal speed
Walking backwards	Walking backwards for 50 m
Running	Jogging 100 meters
Step walking	Moving the feet alternately in the rhythm of a marching step without advancing
Jumping	Jumping for 45s without advancing
Upstairs	Ascending stairs
Downstairs	Descending stairs
Cycling	Cycling with a real bike

human activity categories of the SCUT NAA dataset and their definitions. Since the class of cycling has only 30 samples, we did not include this class in the evaluations. In our experiments, we chose the data collected from accelerometers placed in trousers pocket (TP), randomly selected $p = 20$ and $p = 30$ samples per human activity category for training, and used the remaining samples for testing. In the training stage, we used the training set to learn the orthogonal projection matrix. In the test stage, we project the test set to the low-dimensional space using the projection matrix.

Our NMHA dataset for performance evaluation is created as follows. To collect samples of human activity, we developed an application for Android-powered smart phones. We selected ten typical human activities which are walking, step walking, walking quickly, walking upstairs and walking downstairs, jumping, running, falling, sit to stand, and stand to sit. We collected the raw acceleration data from 61 subjects and there are two positions to fix mobile, such as waist pack (WP), TP, and thus, there are in total two-human activity subdatasets, each of which contains 610 samples.

To inspect the robustness of the proposed scheme, we conduct experiments on each of the two human activity subdatasets and the combination of them. For convenience, we define as follows.

- 1) *NMHA-TP*: Feature set extracted from the acceleration signals collected from TP.
- 2) *NMHA-WP*: Feature set extracted from the acceleration signals collected from WP.
- 3) *NMHA-TP-WP*: The combination of *NMHA-TP* and *NMHA-WP*.

In our experiments on *NMHA-TP* and *NMHA-WP*, we randomly selected $p = 20$ and $p = 30$ samples per human activity category for training, while the remaining samples were used as the test set. In our experiments on *NMHA-TP-WP*, we randomly selected $p = 60$ samples per human activity category for training, while the remaining samples were used as the test set.

Example acceleration signals and activity images are shown in Fig. 2. The horizontal axis represents the sampling points and the sampling interval is 0.01 s. The vertical axis represents the acceleration attributes. For example, $y = 125$ means gravity = $0g$, and each 26 units means $1g$,

where $g = 9.8 \text{ m/s}^2$.

B. Frequency Domain Information

After obtaining the raw acceleration signals from the smart phone, robust feature descriptors are extracted according to the following steps. First, corresponding to each axis (including x , y , and z), we extracted FFT coefficients using a 256-point window. There is a 128-point overlap between consecutive windows. Second, we only retained the first 24 FFT coefficients for each sliding window. Finally, all the FFT coefficients obtained in the second step were concatenated as the frequency domain information for representing a human activity video. The raw acceleration signal of each axis contains 1536 sample points. Thus, we can obtain 11 sampling windows and $11 \times 24 \times 3 = 792$ dimensional coefficients. Note that it is improper to simply use the original signals for classification, because of the alignment problem. Another way to handle this problem is adopting the bag of words model in the FFT domain. Details about the frequency domain information of signals recorded in the SCUT NAA dataset are given in [50].

C. Baselines and Performance Evaluation

In this paper, we compare the effectiveness of the proposed EMRP with five representative algorithms which are PCA [23], LDA [15], [17], GMSS [41], LSDA [9], and MFA [48]. These methods are state-of-the-art dimension reduction methods based on spectral geometry that have shown their merits in various practical problems.

PCA is a global unsupervised algorithm which is widely used in the stage of data preprocessing. LDA is a global supervised algorithm. Compared with PCA, LDA significantly improves the performance of the subsequent classification by exploring the sample label information. However, when the subspace dimension is less than $C - 1$ (C is the number of class), LDA is not optimal. Using the general mean concept to replace the arithmetic mean used in LDA, GMSS can significantly reduce this problem and is more robust than LDA [41]. LSDA and MFA are proposed from the view of local geometrical structure and duly explore the supervised information for the subsequent classification.

Before we conduct LDA, GMSS, LSDA, and MFA, PCA is used as a preprocessing step to make these algorithms avoiding the small sample size problem. Note that: 1) the number of dimensions of the original features is much larger than the number of training samples in LDA [17], so PCA was used to retain $N - C$ dimensions for training samples to ensure that within-scatter matrix S_w is nonsingular, where N is the number of training samples and C is the number of classes; 2) according to [48], we retained $N - C$ dimensions to ensure that $X(D^p - W^p)X^T$ in MFA is nonsingular; and 3) to reduce the training time, we also applied the PCA step to retain $N - 1$ dimensions in LSDA and EMRP.

In our EMRP scheme, the hyperparameters include the heat kernel parameter t , the number of closest intraclass samples k_1 and the number of closest interclass samples k_2 . In particular, we created two alignment matrix sets as follows. For the first

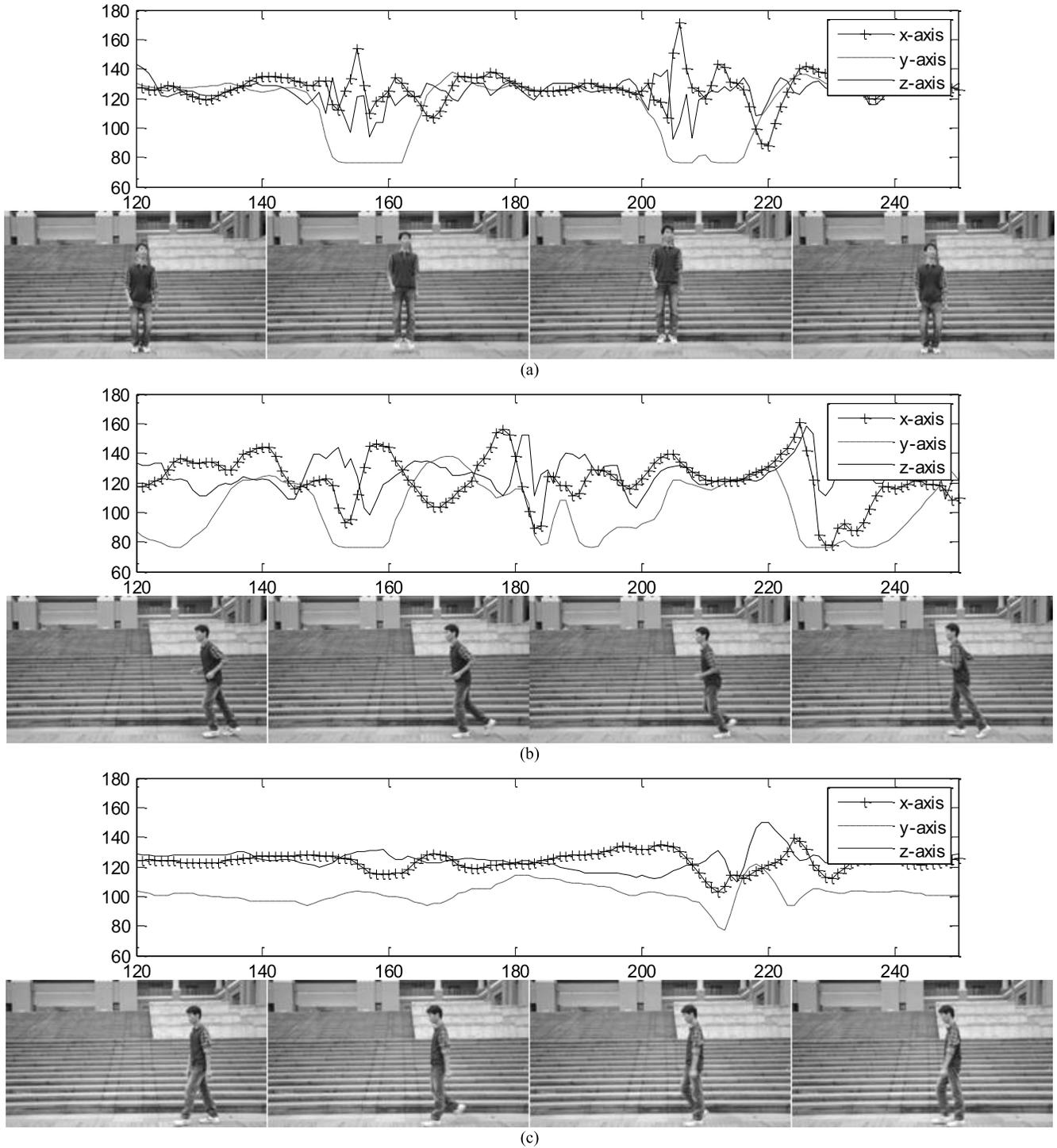


Fig. 2. Example acceleration signals in our NMHA dataset: the dataset contains 10 human activity categoration (walking, step walking, walking quickly, walking upstairs and walking downstairs, jumping, running, falling, sit to stand, and stand to sit). There are three paired samples shown in this figure. For each pair, the acceleration signal is shown on the top (The horizontal axis represents the sampling points and the sampling interval is 0.01 s. The vertical axis represents the acceleration attributes. For example, $y = 125$ means gravity = $0g$, and each 26 units means $1g$, where $g = 9.8 \text{ m/s}^2$.) and the corresponding human activity images is on the bottom. (a) Jumping. (b) Running. (c) Walking.

set, the candidates are

$$t = \left\{ \frac{\theta}{30}, \frac{\theta}{20}, \frac{\theta}{10}, \frac{\theta}{5}, \theta, 5\theta, 10\theta, 20\theta, 30\theta \right\}$$

$k_1 = 5$, and $k_2 = 5$, which led to nine candidate alignment matrices. For the second set, the candidates

are

$$t = \left\{ \frac{\theta}{30}, \frac{\theta}{20}, \frac{\theta}{10}, \frac{\theta}{5}, \theta, 5\theta, 10\theta, 20\theta, 30\theta \right\}$$

$k_1 = \{3, 5, 7\}$, and $k_2 = \{3, 5, 7\}$, which led to totally 81 candidate alignment matrices. Note that θ can be estimated by the inverse of $1/n^2 \sum_{i,j=1}^n \|x_i - x_j\|^2$. We varied all

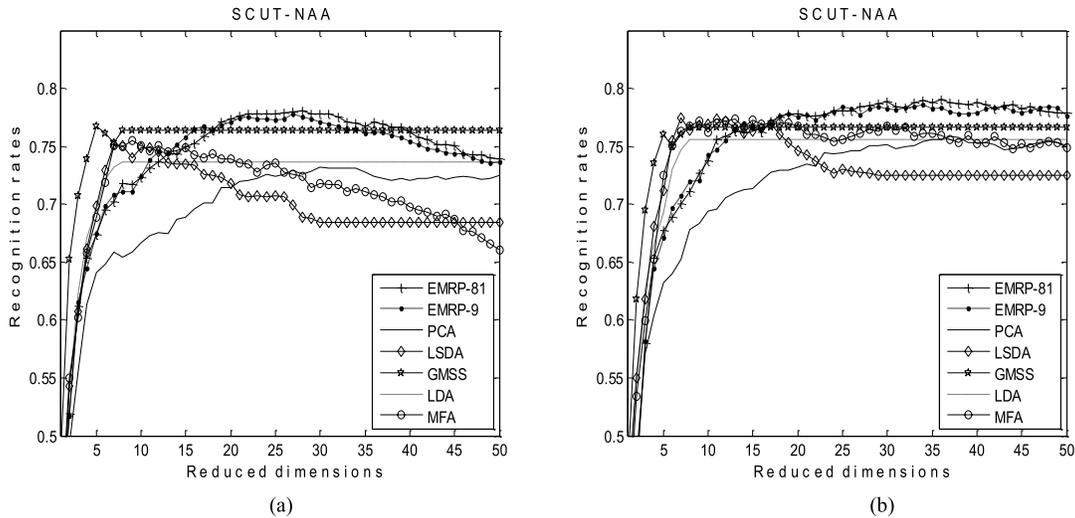


Fig. 3. We compare EMRP with PCA, LSDA, MFA, GMSS, and LDA on the SCUT NAA dataset. The NN classifier is used for recognition. In each subfigure, the x -coordinate is the number of the dimension of all the algorithms on the test set and the y -coordinate is the average recognition. (a) $p = 20$. (b) $p = 30$.

TABLE III
BEST AVERAGE RECOGNITION RATES OF SIX ALGORITHMS ON THE SCUT NAA DATASET

Classes	$p = 20$						$p = 30$							
	PCA	LDA	LSDA	MFA	GMSS	EMRP-9	EMRP-81	PCA	LDA	LSDA	MFA	GMSS	EMRP-9	EMRP-81
Jumping	0.825	0.863	0.871	0.854	0.867	0.842	0.842	0.836	0.864	0.857	0.843	0.850	0.829	0.829
Walking	0.504	0.508	0.529	0.558	0.567	0.608	0.592	0.507	0.493	0.493	0.514	0.514	0.550	0.564
Walking Backwards	0.639	0.617	0.630	0.604	0.709	0.696	0.709	0.623	0.677	0.677	0.615	0.746	0.723	0.723
Walking Quickly	0.575	0.663	0.683	0.667	0.638	0.758	0.763	0.693	0.686	0.729	0.736	0.729	0.829	0.821
Running	0.842	0.842	0.846	0.879	0.867	0.963	0.963	0.864	0.850	0.871	0.921	0.879	0.964	0.971
Relaxing	1.000	1.000	0.996	1.000	1.000	0.992	0.992	1.000	1.000	1.000	1.000	1.000	0.993	0.993
Downstairs	0.858	0.746	0.792	0.821	0.846	0.725	0.738	0.879	0.843	0.914	0.879	0.800	0.807	0.807
Upstairs	0.704	0.588	0.588	0.629	0.533	0.604	0.604	0.700	0.536	0.593	0.614	0.464	0.614	0.614
Step walking	0.633	0.796	0.821	0.779	0.879	0.813	0.821	0.714	0.850	0.829	0.829	0.914	0.757	0.786
Average accuracy	0.732	0.736	0.751	0.755	0.767	0.778	0.780	0.758	0.756	0.774	0.774	0.766	0.785	0.791
	(30)	(8)	(7)	(9)	(5)	(27)	(28)	(37)	(8)	(7)	(13)	(8)	(34)	(36)

(The number in the parentheses is the reduced dimensions.)

hyperparameters in a larger range to obtain a robust approximation. According to the different hyperparameter candidate sets, i.e., the first and the second sets, we named the proposed algorithms EMRP-9 and EMRP-81, respectively.

In addition, considering to unify the classification stage, k NN classifier ($k = 1$) was used to cooperate with various spectral geometry algorithms. These trials were independently repeated for 10 times and the average accuracy for each class was reported for performance evaluation.

D. Experimental Results and Analysis

In Fig. 3, we compare the proposed EMRP with PCA, LDA, GMSS, LSDA, and MFA on the SCUT NAA dataset. In each subfigure, the x -coordinate is the number of dimensions of subspace and the y -coordinate is the average accuracy. In Table III, we reported the best average recognition rates of six algorithms and the corresponding dimension on the SCUT NAA dataset.

In Fig. 4, we compare the proposed EMRP with PCA, LDA, GMSS, LSDA, and MFA on the NMHA-TP dataset. In Table IV, we reported the best average recognition rates of six algorithms and the corresponding dimension on the NMHA-TP dataset.

In Fig. 5, we compare the proposed EMRP with PCA, LDA, GMSS, LSDA, and MFA on the NMHA-WP dataset. In Table V, we reported the best average recognition rates of six algorithms and the corresponding dimension on the NMHA-WP dataset.

In Fig. 6, we compare the proposed EMRP with PCA, LDA, GMSS, LSDA, and MFA on the NMHA-TP-WP dataset. In Table VI, we reported the best average recognition rates of six algorithms and the corresponding dimension on the NMHA-TP-WP dataset. In addition, under the same experimental setting, the NN classifier has been replaced with the SVM classifier (using the RBF kernel). We can observe that the best accuracy is achieved by EMRP-81

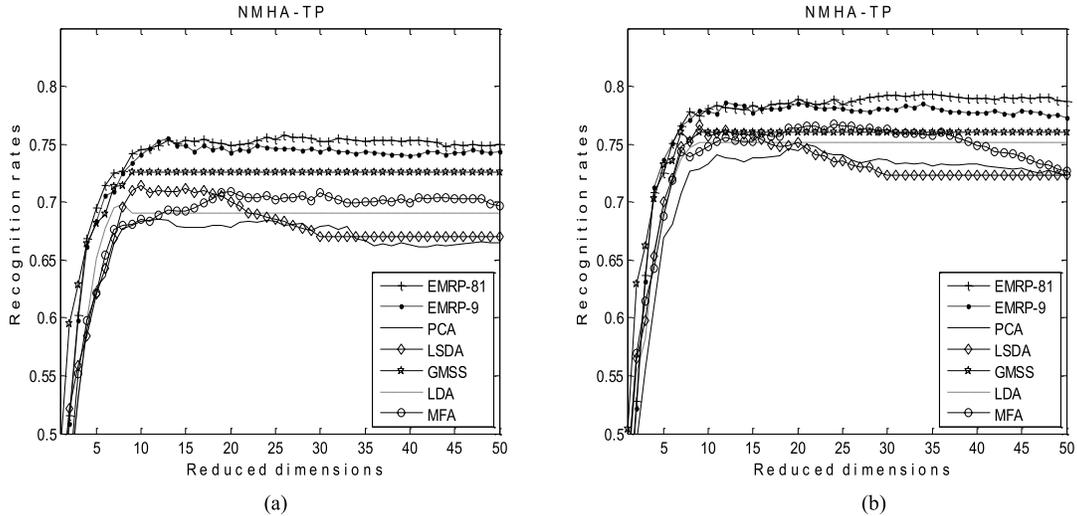


Fig. 4. We compare EMRP with PCA, LSDA, MFA, GMSS, and LDA on the NMHA-TP dataset. The NN classifier is used for recognition. In each subfigure, the x -coordinate is the number of the dimension of all the algorithms on the test set and the y -coordinate is the average recognition. (a) $p = 20$. (b) $p = 30$.

TABLE IV
BEST AVERAGE RECOGNITION RATES OF SIX ALGORITHMS ON THE NMHA-TP DATASET

Classes	$p = 20$							$p = 30$						
	PCA	LDA	LSDA	MFA	GMSS	EMRP-9	EMRP-81	PCA	LDA	LSDA	MFA	GMSS	EMRP-9	EMRP-81
Falling	0.717	0.583	0.732	0.637	0.629	0.761	0.732	0.758	0.671	0.729	0.723	0.739	0.745	0.806
Jumping	0.805	0.754	0.663	0.629	0.717	0.824	0.827	0.845	0.794	0.765	0.752	0.777	0.842	0.858
Running	0.824	0.785	0.715	0.741	0.771	0.868	0.880	0.881	0.835	0.839	0.810	0.819	0.939	0.948
sit to stand	0.500	0.515	0.620	0.551	0.605	0.507	0.534	0.587	0.574	0.635	0.655	0.629	0.597	0.552
stand to sit	0.534	0.500	0.700	0.654	0.732	0.629	0.663	0.600	0.603	0.648	0.674	0.700	0.697	0.742
Walking	0.571	0.722	0.734	0.790	0.793	0.700	0.739	0.658	0.768	0.758	0.784	0.806	0.716	0.732
Step walking	0.717	0.839	0.802	0.868	0.844	0.839	0.822	0.819	0.919	0.919	0.926	0.871	0.871	0.871
Walking quickly	0.588	0.622	0.559	0.578	0.529	0.637	0.651	0.587	0.658	0.671	0.671	0.600	0.661	0.700
Downstairs	0.805	0.893	0.783	0.817	0.815	0.912	0.873	0.919	0.839	0.865	0.839	0.832	0.945	0.903
Upstairs	0.790	0.776	0.832	0.822	0.824	0.871	0.854	0.839	0.852	0.832	0.839	0.832	0.845	0.819
Average accuracy	0.685	0.699	0.714	0.710	0.726	0.754	0.758	0.749	0.751	0.766	0.767	0.761	0.785	0.793
	(12)	(8)	(10)	(20)	(9)	(13)	(26)	(21)	(9)	(9)	(24)	(9)	(12)	(35)

(The number in the parentheses is the reduced dimensions.)

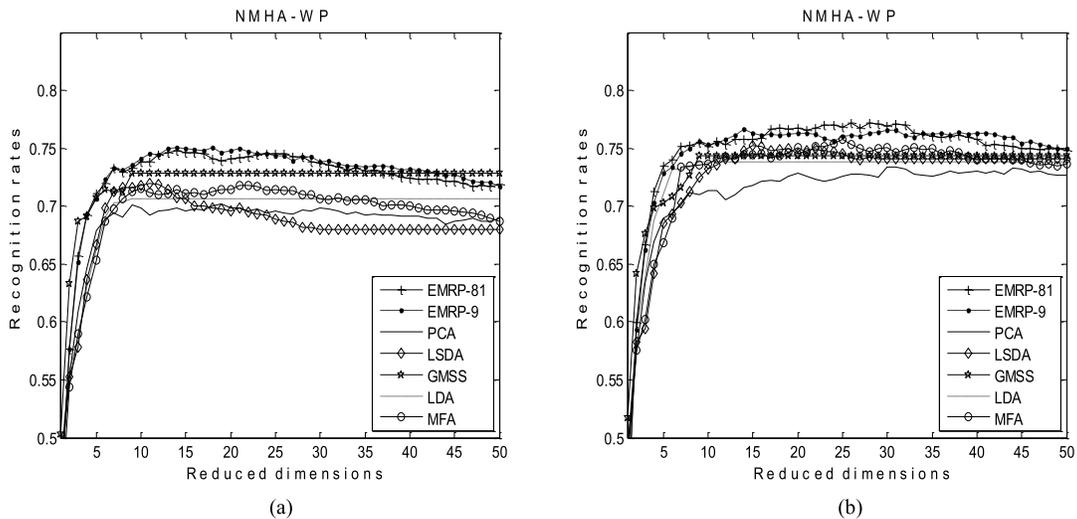


Fig. 5. We compare EMRP with PCA, LSDA, MFA, GMSS, and LDA on the NMHA-WP dataset. The NN classifier is used for recognition. In each subfigure, the x -coordinate is the number of the dimension of all the algorithms on the test set and the y -coordinate is the average recognition. (a) $p = 20$. (b) $p = 30$.

in Table VII. It shows that the SVM classifier can further improve the accuracy of accelerometer-based human activity recognition.

Fig. 7 shows the classification confusion matrix of EMRP for one test split of the NMHA-TP-WP dataset. All correct predications are located in the diagonal of the matrix. It is

TABLE V
BEST AVERAGE RECOGNITION RATES OF SIX ALGORITHMS ON THE NMHA-WP DATASET

Classes	$p = 20$							$p = 30$						
	PCA	LDA	LSDA	MFA	GMSS	EMRP-9	EMRP-81	PCA	LDA	LSDA	MFA	GMSS	EMRP-9	EMRP-81
Falling	0.805	0.776	0.805	0.781	0.822	0.873	0.861	0.874	0.897	0.894	0.800	0.855	0.897	0.897
Jumping	0.658	0.698	0.695	0.627	0.627	0.739	0.746	0.729	0.761	0.752	0.694	0.687	0.787	0.803
Running	0.805	0.849	0.849	0.849	0.817	0.905	0.907	0.829	0.858	0.865	0.904	0.871	0.939	0.939
sit to stand	0.590	0.624	0.617	0.640	0.680	0.688	0.688	0.635	0.639	0.639	0.700	0.623	0.668	0.690
stand to sit	0.614	0.605	0.615	0.598	0.668	0.588	0.590	0.619	0.574	0.606	0.678	0.671	0.642	0.606
Walking	0.563	0.568	0.600	0.727	0.739	0.607	0.602	0.622	0.597	0.594	0.733	0.768	0.629	0.668
Step walking	0.961	0.927	0.944	0.920	0.929	0.966	0.973	0.961	0.913	0.965	0.942	0.948	0.958	0.971
Walking quickly	0.561	0.595	0.602	0.620	0.634	0.615	0.627	0.600	0.665	0.694	0.636	0.687	0.700	0.739
Downstairs	0.739	0.751	0.783	0.674	0.644	0.783	0.785	0.738	0.748	0.742	0.762	0.597	0.732	0.690
Upstairs	0.724	0.668	0.683	0.747	0.720	0.715	0.727	0.735	0.729	0.777	0.733	0.726	0.710	0.716
Average accuracy	0.702 (19)	0.706 (9)	0.719 (11)	0.718 (22)	0.728 (9)	0.747 (14)	0.751 (14)	0.734 (31)	0.738 (9)	0.753 (15)	0.758 (25)	0.743 (9)	0.766 (14)	0.772 (28)

(The number in the parentheses is the reduced dimensions.)

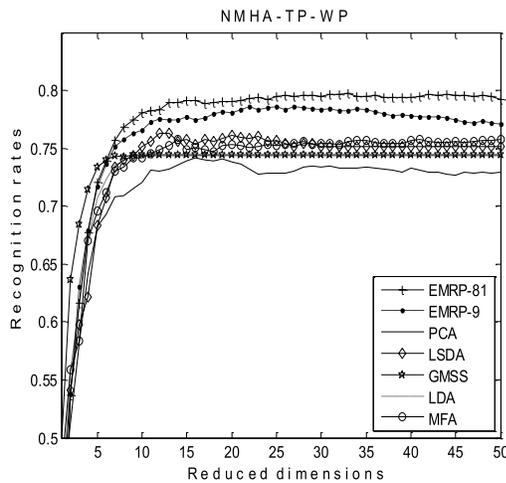


Fig. 6. We compare EMRP with PCA, LSDA, MFA, GMSS, and LDA on the NMHA-TP-WP dataset.

convenient to inspect the picture for classification errors, as they are nonzero values listed outside the diagonal. It is not surprising that confusions occur between sit to stand and stand to sit, because mobiles are not fixed in the body of volunteers and the mobiles position directly impacts the recognition accuracy of these two human activities. From the confusion matrix, we can also find that falling is confused with stand to sit, because there are very similar.

The main observations from the experiment of EMRP can be summarized as follows.

- 1) EMRP, LSDA, and MFA are promising solutions, because they all consider the geometric information in a local patch intuitively. EMRP outperforms others, because it aims to find the balance of the intraclass nearest neighborhood ranks and the interclass nearest neighborhood ranks, which is important for improving the accuracy of human behavior categorization.
- 2) EMRP-81 significantly improves the recognition accuracy and obtains the best average recognition rates in our human activity recognition experiments.

TABLE VI

BEST AVERAGE RECOGNITION RATES OF SIX ALGORITHMS ON THE NMHA-TP-WP DATASET

Classes	$p = 60$						
	PCA	LDA	LSDA	MFA	GMSS	EMRP-9	EMRP-81
Falling	0.770	0.706	0.771	0.671	0.692	0.739	0.785
Jumping	0.849	0.860	0.850	0.795	0.821	0.910	0.911
Running	0.901	0.876	0.868	0.898	0.898	0.953	0.955
sit to stand	0.588	0.584	0.558	0.611	0.485	0.606	0.585
stand to sit	0.619	0.608	0.658	0.671	0.440	0.702	0.695
Walking	0.682	0.663	0.706	0.763	0.768	0.768	0.792
Step walking	0.859	0.882	0.965	0.932	0.950	0.924	0.937
Walking quickly	0.645	0.663	0.687	0.711	0.718	0.719	0.731
Downstairs	0.790	0.840	0.785	0.792	0.879	0.823	0.834
Upstairs	0.719	0.748	0.782	0.732	0.792	0.718	0.748
Average accuracy	0.742 (16)	0.743 (9)	0.763 (12)	0.758 (50)	0.744 (9)	0.786 (25)	0.797 (33)

(The number in the parentheses is the reduced dimensions.)

TABLE VII

BEST AVERAGE RECOGNITION RATES USING SVM CLASSIFIER

	PCA	LDA	LSDA	MFA	GMSS	EMRP-9	EMRP-81
Average accuracy	0.760	0.763	0.785	0.778	0.772	0.802	0.813

This demonstrates the robustness and effectiveness of EMRP for human behavior categorization task. This justifies that EMRP-based algorithms can learn a suitable linear combination of the alignment matrices for obtaining better system performance.

- 3) Figs. 3 and 5 demonstrate that the recognition rates of the proposed EMRP decrease with the increase of reduced dimensions, after the optimal value of dimen-

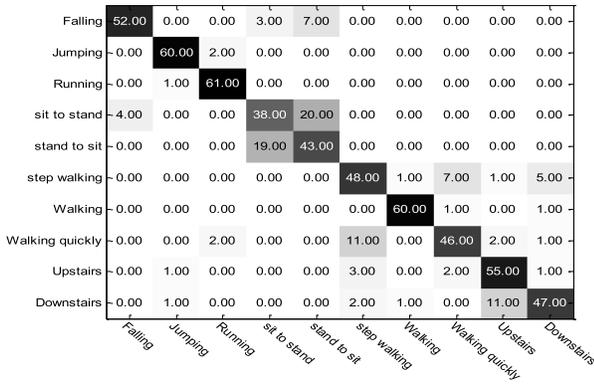


Fig. 7. EMRP classification confusion matrix for one test split on the NMHA-TP-WP dataset. The NN classifier is used for recognition.

sion is reached. This is because the number of training samples is much smaller than the dimension of feature vectors.

V. CONCLUSION

In this paper, a new spectral geometry algorithm termed EMRP is proposed for acceleration-based human activity recognition. By considering the concentration of the measure phenomenon, EMRP preserves the rank order information in local patches. By maximizing the margin between samples of different classes, EMRP retains the most discriminative information. Especially, EMRP finds the optimal linear combination of the alignment matrices to approximate the intrinsic manifold lied in the data.

Compared to the classical spectral geometry algorithms, such as PCA, LDA, GMSS, LSDA, and MFA, EMRP shows many attractive and competitive properties to measure the similarity between different acceleration signals. In the acceleration-based human activity recognition application, the newly proposed nonmetric scheme based on EMRP is superior to the above methods in terms of recognition accuracy.

REFERENCES

- [1] A. G. Bonomi, A. H. Goris, B. Yin, and K. R. Westertep, "Detection of type, duration, and intensity of physical activity using an accelerometer," *Med. Sci. Sports Exerc.*, vol. 41, no. 9, pp. 1770–1777, 2009.
- [2] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Proc. 2nd Int. Conf. Pervasive Comput.*, 2004, pp. 1–17.
- [3] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
- [4] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing System*, vol. 14. Cambridge, MA, USA: MIT Press, 2002, pp. 585–591.
- [5] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [6] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.
- [7] S. R. Bulò and M. Pelillo, "A game-theoretic approach to hypergraph clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1312–1327, Jun. 2013.
- [8] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
- [9] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," in *Proc. 20th IJCAI*, 2007, pp. 708–713.
- [10] A. Carli, U. Castellani, M. Bicego, and V. Murino, "Dissimilarity-based representation for local parts," in *Proc. IEEE 2nd Int. Workshop Cognit. Inf. Process.*, Jun. 2010, pp. 299–303.
- [11] Y. Chen, J. Zhang, D. Cai, W. Liu, and X. He, "Nonnegative local coordinate factorization for image representation," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 969–979, Mar. 2013.
- [12] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," *Neural Netw.*, vol. 32, pp. 333–338, Aug. 2012.
- [13] P. Demartines and J. Hérault, "Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 148–154, Jan. 1997.
- [14] D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," in *Math Challenges Lecture*. Los Angeles, CA, USA: AMS, 2000.
- [15] F. Dufrenois and J. C. Noyer, "Formulating robust linear regression estimation as a one-class LDA criterion: Discriminative hat matrix," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 2, pp. 262–273, Feb. 2013.
- [16] M. Ermers, J. Pärkkä, J. Mäntyjärvi, and I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 1, pp. 20–26, Jan. 2008.
- [17] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [18] B. Geng, D. Tao, C. Xu, L. Yang, and X.-S. Hua, "Ensemble manifold regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1227–1233, Jun. 2012.
- [19] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2030–2048, Jul. 2011.
- [20] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Non-negative patch alignment framework," *IEEE Trans. Neural Netw.*, vol. 22, no. 8, pp. 1218–1230, Aug. 2011.
- [21] L. Han, F. Escolano, E. R. Hancock, and R. C. Wilson, "Graph characterizations from von Neumann entropy," *Pattern Recognit. Lett.*, vol. 33, no. 15, pp. 1958–1967, 2012.
- [22] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing System*, vol. 16. Cambridge, MA, USA: MIT Press, 2004.
- [23] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Edu. Psychol.*, vol. 24, no. 6, pp. 417–441, 1933.
- [24] A. Iosifidis, A. Tefas, and I. Pitas, "View-invariant action recognition based on artificial neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 3, pp. 412–424, Mar. 2012.
- [25] D. W. Jacobs, D. Weinshall, and Y. Gdalyahu, "Classification with nonmetric distances: Image retrieval and class representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 6, pp. 583–600, Jun. 2000.
- [26] 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.
- [27] D. Keysers, T. Deselaers, C. Gollan, and H. Ney, "Deformation models for image recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1422–1435, Aug. 2007.
- [28] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explorations Newslett.*, vol. 12, no. 2, pp. 74–82, Dec. 2010.
- [29] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 2169–2178.
- [30] S. Lespinats, M. Verleysen, A. Giron, and B. Fertil, "DD-HDS: A method for visualization and exploration of high-dimensional data," *IEEE Trans. Neural Netw.*, vol. 18, no. 5, pp. 1265–1279, Sep. 2007.
- [31] X. Long, B. Yin, and R. M. Aarts, "Single-accelerometer based daily physical activity classification," in *Proc. Annu. Int. Conf. IEEE EMBC*, Minneapolis, MN, USA, Sep. 2009, pp. 6107–6110.
- [32] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

- [33] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [34] V. Osmani, S. Balasubramaniam, and D. Botvich, "Human activity recognition in pervasive health-care: Supporting efficient remote collaboration," *J. Netw. Comput. Appl.*, vol. 31, no. 4, pp. 628–655, 2008.
- [35] Y. Pang, Z. Ji, P. Jing, and X. Li, "Ranking graph embedding for learning to rerank," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 8, pp. 1292–1303, Aug. 2013.
- [36] T. K. Paul and T. Ogunfunmi, "Study of the convergence behavior of the complex kernel least mean square algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 9, pp. 1349–1363, Sep. 2013.
- [37] N. Rebagliati, A. Sole-Ribalta, M. Pelillo, and F. Serratos, "Computing the graph edit distance using dominant sets," in *Proc. IEEE 1st Int. Conf. Pattern Recognit.*, Nov. 2012, pp. 1080–1083.
- [38] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [39] L. Shao, L. Ji, Y. Liu, and J. Zhang, "Human action segmentation and recognition via motion and shape analysis," *Pattern Recognit. Lett.*, vol. 33, no. 4, pp. 438–445, Mar. 2012.
- [40] D. Tao, L. Jin, Y. Wang, and X. Li, "Rank preserving discriminant analysis for human behavior recognition on wireless sensor networks," *IEEE Trans. Ind. Informat.*, vol. 10, no. 1, pp. 813–823, Feb. 2014.
- [41] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.
- [42] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [43] M. Tentori and J. Favela, "Activity-aware computing for healthcare," *IEEE Pervasive Comput.*, vol. 7, no. 2, pp. 51–57, Apr./Jun. 2008.
- [44] D. Tosato, M. Spera, M. Cristani, and V. Murino, "Characterizing humans on Riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1972–1984, Aug. 2013.
- [45] W. Ugulino, D. Cardador, K. Vega, E. Velloso, R. Milidiú, and H. Fuks, "Wearable computing: Accelerometers' data classification of body postures and movements," in *Proc. 21th Brazilian Symp. Artif. Intell. Adv. Artif. Intell. (SBIA)*, 2012, pp. 52–61.
- [46] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3360–3367.
- [47] X. Wang, D. Rosenblum, and Y. Wang, "Context-aware mobile music recommendation for daily activities," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 99–108.
- [48] 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 retrieval," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2811–2821, Nov. 2007.
- [49] W. Xu, E. R. Hancock, and R. C. Wilson, "Ricci flow embedding for rectifying non-Euclidean dissimilarity data," *Pattern Recognit.*, vol. 47, no. 11, pp. 3709–3725, 2014.
- [50] Y. Xue and L. Jin, "A naturalistic 3D acceleration-based activity dataset & benchmark evaluations," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2010, pp. 4081–4085.
- [51] P. Yan, W. Zhang, B. Turkbey, P. L. Choyke, and X. Li, "Global structure constrained local shape prior estimation for medical image segmentation," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 1017–1026, 2013.
- [52] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, Sep. 2009.
- [53] Z. Zhang, P. Ren, and E. R. Hancock, "Unsupervised feature selection via hypergraph embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 39.1–39.11.



Dapeng Tao received the B.E. degree from Northwestern Polytechnical University, Xi'an, China, and the Ph.D. degree from the South China University of Technology, Guangzhou, China.

He has authored or co-authored over 30 scientific articles. His research interests include machine learning, computer vision, and cloud computing.

Dr. Tao has served over 10 international journals, including the *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, the *IEEE TRANSACTIONS ON MULTIMEDIA*, the *IEEE SIGNAL PROCESSING LETTERS*, and *PLOS ONE*.



Lianwen Jin (M'09) received the B.S. degree from the University of Science and Technology of China, Hefei, China, in 1991, and the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 1996.

He is currently a Professor with the School of Electronic and Information Engineering, South China University of Technology. He has authored over 100 scientific papers. His current research interests include image processing, handwriting analysis and recognition, machine learning, cloud computing,

and intelligent systems.

Dr. Jin is a member of the IEEE Signal Processing Society, the IEEE Communication Society, the IEEE Computer Society, the China Image and Graphics Society, and the Cloud Computing Experts Committee of the China Institute of Communications. He was a recipient of the New Century Excellent Talent Program Award of the Ministry of Education in 2006 and the Guangdong Pearl River Distinguished Professor Award in 2011. He served as the Program Committee Member for a number of international conferences, including the International Conference on Machine Learning and Cybernetics from 2007 to 2011, the International Conference on Frontiers in Handwriting Recognition 2008–2012, the International Conference on Document Analysis and Recognition in 2009 and 2013, the International Conference on Pattern Recognition in 2010 and 2012, and the International Conference on Machine Learning and Applications in 2012.

Yuan Yuan (M'05–SM'09) is currently a Full Professor with the Chinese Academy of Sciences, Beijing, China. Her current research interests include visual information processing and image/video content analysis. She has authored over 100 papers, including over 70 in reputable journals, such as the *IEEE TRANSACTIONS* and *Pattern Recognition*.

Dr. Yuan also authored conference papers in the Computer Vision and Pattern Recognition Conference, the British Machine Vision Conference, the International Conference on Image Processing, and the International Conference on Acoustics, Speech and Signal Processing.



Yang Xue received the B.S. degree from Northeastern University, Shenyang, China, and the Ph.D. in electronic and information engineering from the South China University of Technology (SCUT), Guangzhou, China.

She is currently an Associate Professor with the School of Electronic and Information Engineering, SCUT. Her current research interests include wearable computing, context-aware computing, accelerometer-based human activity recognition, and human-computer interaction intelligent systems.