3D Ear Identification Using Block-Wise Statistics-Based Features and LC-KSVD

Lin Zhang, *Senior Member, IEEE*, Lida Li, Hongyu Li, and Meng Yang, *Member, IEEE*

***Abstract*—Biometrics authentication has been corroborated to be an effective method for recognizing a person's identity with high confidence. In this field, the use of three-dimensional (3D) ear shape is a recent trend. As a biometric identifier, the ear has several inherent merits. However, although a great deal of efforts have been devoted, there is still large room for improvement in developing a highly effective and efficient 3D ear identification approach. In this paper, we attempt to fill this gap to some extent by proposing a novel 3D ear classification scheme that makes use of the label consistent K-SVD (LC-KSVD) framework. As an effective supervised dictionary learning algorithm, LC-KSVD learns a single compact discriminative dictionary for sparse coding and a multi-class linear classifier simultaneously. To use the LC-KSVD framework, one key issue is how to extract feature vectors from 3D ear scans. To this end, we propose a block-wise statistics-based feature extraction scheme. Specifically, we divide a 3D ear region of interest into uniform blocks and extract a histogram of surface types from each block; histograms from all blocks are then concatenated to form the desired feature vector. Feature vectors extracted in this way are highly discriminative and are robust to mere misalignment between samples. Experiments demonstrate that our approach can achieve better recognition accuracy than the other state-of-the-art methods. More importantly, its computational complexity is extremely low, making it quite suitable for the large-scale identification applications. MATLAB source codes are publicly online available at http://sse.tongji.edu.cn/linzhang/LCKSVDEar/LCKSVDEar.htm.**

***Index Terms*—3D ear, dictionary learning, label consistent K-SVD (LC-KSVD), sparse coding, surface types.**

I. Introduction

THE need for reliable user authentication techniques are significantly increased in the wake of heightened concerns about security, and rapid advancement in networking, communication and mobility [1]. Propelled by the requirements of numerous applications, such as access control, aviation security, or e-banking, automatically recognizing the identity of a person with high confidence has become a topic of intense study. To solve such a problem, biometrics based methods, which use unique physical or behavioral characteristics of human beings, are drawing increasing attention recently because of their high accuracy and robustness. In the past several decades or so, researchers have exhaustively investigated a number of different biometric identifiers, such as fingerprint [2], [3], face [4]–[6], iris [7], [8], palmprint [9]–[11], hand geometry [12], gait [13], finger-knuckle-print [14], [15], etc.

Among many biometric identifiers, the ear has recently received significant attention due to its non-intrusiveness and ease of data collection. As a biometric identifier, ear is appealing and has some desirable properties. For example, compared with face, ear recognition is less likely to be affected by various facial expressions. Ear has a rich structure and a distinct shape which remains unchanged from 8 to 70 years of age as determined by Iannarelli through a study of 10 000 ears [16]. As pointed out by Chang *et al*. [17], the recognition using two-dimensional (2D) ear images has a comparable discriminative power compared with the recognition using 2D face images.

Much progress has been made in the fields relevant to ear recognition in recent years. Ear recognition problems can be roughly classified as 2D, 3D, and multimodal 2D plus 3D, according to the types of input data. Most studies in this field in the early stage exploited only 2D profile images and representative works can be found in [17]–[22]. However, it has been observed that variations between the images of the same ear due to changes of illumination or viewing direction are often larger than those caused by changes in ear identity. The introduction of the 3D modality mitigates some of these challenges by introducing a depth dimension that is invariant to both lighting conditions and head pose. With the development and the popularization of the 3D sensing technology, there is a rising trend to use 3D sensors instead of 2D cameras in the field of ear recognition. Compared with its 2D counterpart, 3D data contains more abundant information about the ear shape and is more robust to illumination variations and occlusions. Yan and Bowyer found that ear matching based on 3D data could achieve a higher recognition accuracy than that making use of the corresponding 2D images [23].

However, how to devise a highly effective and efficient 3D ear identification approach is still an open issue and in this paper we try to solve this problem to some extent. The remainder of this paper is organized as follows. Section II introduces the related works and our contributions. Section III presents our

Manuscript received December 09, 2015; revised April 09, 2016; accepted May 03, 2016. Date of publication May 10, 2016; date of current version July 15, 2016. This work was supported in part by the Natural Science Foundation of China under Grant 61201394 and Grant 61402289, in part by the Shanghai Pujiang Program under Grant 14PJ1408100, and in part by the CCF-Tencent Open Research Fund under Grant CCF-Tencent RAGR20150112. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Cha Zhang.

L. Zhang is with the School of Software Engineering, Tongji University, Shanghai 201804, China, the Shenzhen Institute of Future Media Technology, Shenzhen 518055, China, and also with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: cslinzhang@tongji.edu.cn).

L. Li is with the School of Software Engineering, Tongji University, Shanghai 201804, China (e-mail: lld533@hotmail.com).

H. Li is with the School of Software Engineering, Tongji University, Shanghai 201804, China, and also with the Xiamen Macrovis Technology Company, Ltd., Xiamen 361008, China (e-mail: hyli@tongji.edu.cn).

M. Yang is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: yang.meng@szu.edu.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/TMM.2016.2566578

proposed block-wise statistics based feature extraction scheme. Section IV presents our proposed 3D ear classification scheme in detail. Experimental results are presented in Section V. Finally, Section VI concludes the paper.

II. Related Works and Our Contributions

A. 3D Ear Detection and Classification

To construct a real 3D ear based personal authentication system, there are two key components, ear region detection and ear matching. In the literature, several different schemes have been proposed for 3D ear region detection. Among them, some are totally based on 3D range data, e.g. [24]–[30], while others are based on multi-modal co-registered 2D plus 3D data [31]–[34]. In [24], Chen and Bhanu proposed a two-step approach to detect the ear region, which includes model template building and online detection. The model template is obtained by averaging the shape index histograms of multiple ear samples and the online detection includes four steps, namely, step edge detection and thresholding, image dilation, connected-component labeling, and template matching. In their later work [25], Chen and Bhanu represented an ear shape model by a set of discrete 3D vertices on the ear helix and anti-helix parts and aligned the model with the range images to detect the ear regions. In [26]–[29], Passalis *et al.* proposed a generic annotated ear model to register and fit each 3D ear and then a compact biometric signature was extracted. In [30], Zhang *et al.* proposed an ear contour alignment based ear detection method. With their method, a range image is at first transformed to a canonical frame by aligning it with an ear contour template created offline and then the ear region is extracted accordingly. In [31], the ear region was initially located by taking a predefined sector from the nose tip. The non-ear portion was then cropped out from that sector using a skin detection algorithm and the ear pit was detected using Gaussian smoothing and curvature estimation algorithms. Then, an active contour algorithm was exploited to extract the ear contour. In [32], [33], ear regions are detected from 2D profile images by training an AdaBoost classifier and then the corresponding 3D ear data is extracted from the co-registered 3D profile data. In [34], Chen and Bhanu also resorted to both color and range images to extract ear regions. They used a reference ear shape model based on the helix and anti-helix curves and the global-to-local shape registration.

With respect to the 3D ear matching schemes, most of the existing state-of-the-art methods [31]–[37] adopt iterative closest point (ICP) [38] or its variants. While ICP is an appealing approach for the one-to-one verification applications, it is not quite suitable for the one-to-many identification case. Roughly speaking, ICP-based matching is quite time consuming. If there are multiple samples for each subject in the gallery set, to figure out the identity of a given test sample using an ICP-based matching method, it would be necessary to match the test sample to all the samples in the gallery set one by one. Such a brute-force searching strategy is obviously not quite computationally efficient, especially when the size of the gallery set is extremely large. Therefore, ICP-based methods are not appropriate for dealing with large-scale identification applications. In [30], Zhang *et al.* tried to solve this problem by using the sparse representation based classification framework. In their method, feature vectors are extracted from ear samples in the gallery set and they form an overcomplete dictionary $\mathbf{A}$. It implies that if sufficient training samples are available from each class, it will be possible to represent the test sample as a linear combination of just those training samples in $\mathbf{A}$ from the same class. When a test sample is presented, its feature vector $\mathbf{y}$ is extracted at first and then $\mathbf{y}$ is coded over the dictionary $\mathbf{A}$; the identity of $\mathbf{y}$ can be figured out by checking which class leads to the minimum representation error.

For a more comprehensive recent review about ear recognition, readers can refer to [39].

B. Sparse Coding and Dictionary Learning

Since our proposed 3D ear identification approach has a close relationship with sparse coding and dictionary learning, some recent developments in these two fields will be briefly reviewed here.

In recent years, sparse coding has been successfully explored to solve a variety of problems in computer vision and image analysis, e.g. image denoising [40], image restoration [41], [42], object classification [30], [43]–[45], musical signal analysis [46], and blind image quality assessment [47].

With sparse coding, an input signal $\mathbf{y}$ is approximated by a linear combination of a few items from an overcomplete dictionary $\mathbf{A}$ and usually $\mathbf{y}$'s identity can be determined by evaluating which class yields the least reconstruction error. As pointed out in [48], usually a dictionary learned from the training samples can produce better results than the one using off-the-shelf bases that are predefined and do not depend on any specific data, such as Fourier or wavelet bases. In [43], Wright *et al.* employ the entire set of training samples as the dictionary for discriminative sparse coding and they achieve impressive performances for face recognition. However, determining sparse codes from large dictionaries is quite computationally expensive, prohibiting real-time applications. Hence, to scale to large training sets, compact dictionary learning approaches have been developed and several prominent supervised dictionary learning methods will be briefly reviewed here.

In [49] and [50], one dictionary is learned for each class; classification is performed based on the corresponding reconstruction errors. In [51], Ramirez *et al.* learn class specific dictionaries with an incoherence promoting term, which encourages class specific dictionaries to be independent. In [52], Zhang *et al.* wrap the dictionary learning process inside a boosting procedure for learning multiple dictionaries. In [53], multiple dictionaries for visually correlated object categories are learned; a common shared dictionary is used to characterize common visual properties of the group and multiple category-specific dictionaries are used to capture category-specific visual properties. The drawback of learning class specific dictionaries is that dictionary construction during training and class-wise sparse coding during testing are both quite time consuming when the number of classes is large. Some approaches learn a compact dictionary by merging dictionary items from an initially large dictionary. For example in [54], Winn *et al.* propose to merge the visual items by considering the tradeoff between intra-class

compactness and inter-class discrimination power. In [55], the dictionary is learned through merging two items by maximizing the mutual information of class distributions. In [56], Fulkerson *et al.* construct small dictionaries that can maintain the performance of their larger counterparts by using agglomerative information bottleneck [57]. Some recently proposed approaches incorporate discriminative terms into the objective function during training in order to obtain dictionaries having outstanding capability for classification. The approach proposed in [58] iteratively updates the dictionary based on the outcome of a linear classifier. It may be stuck in a local minimum since it alternates between the dictionary construction and the classifier learning. In [59], Yang *et al.* learn a structured dictionary, in which the dictionary atoms have correspondence to the class labels, by employing the Fisher discrimination criterion. Quite recently, Jiang *et al.* proposed a dictionary learning method, namely label consistent K-SVD (LC-KSVD) [60]. In their method, in addition to using class labels of training data, they also associate label information with each dictionary item to enforce discriminability in sparse codes during the dictionary learning process. With LC-KSVD, a single overcomplete dictionary and an optimal linear classifier can be learned simultaneously.

C. Overview of Our Approach

As aforementioned, though 3D ear is an attractive biometric trait, how to construct a highly effective and efficient identification system based on 3D ear is still an open issue. In this paper, we aim to bring some new improvements to this field. It needs to be noted that we only focus on investigating the ear classification methods. For 3D ear region of interest (ROI) extraction, we use the method proposed in [30]. In this paper, we assume that 3D ear ROIs have already been available.

On seeing that the supervised dictionary learning techniques have achieved great success in various different fields, we attempt to adapt them for 3D ear identification. Specifically, our approach is based on LC-KSVD [60], since pleasing results have been reported by using it in several different classification tasks, including face classification, object classification, scene classification, and action classification. With LC-KSVD, in addition to a compact discriminative dictionary, a multiclass linear classifier can also be learned jointly, which makes the classification rather efficient. To our knowledge, our work is the first one introducing supervised dictionary learning techniques into the field of 3D ear identification.

To adapt LC-KSVD for 3D ear identification, how to extract feature vectors to represent 3D ears is a rather critical issue. Since there exists mere misalignments between two ear ROIs, the extracted feature vectors should be robust to small misalignments while maintaining a high discriminative capability. To meet these requirements, we propose a novel block-wise statistics based feature extraction scheme. Specifically, we at first divide a 3D ear ROI into uniform blocks and extract a histogram of surface types (STs) [61] from each block; histograms from all the blocks are then concatenated to form the final feature vector. Experimental results demonstrate that such feature vectors are highly discriminative and are robust to mere misalignment between ear samples.

The effectiveness and the efficiency of our proposed 3D ear identification scheme has been corroborated by extensive experiments conducted on the benchmark datasets. To make the results fully reproducible, MATLAB source codes of our approach have been made public online available at http://sse.tongji.edu.cn/linzhang/LCKSVDEar/LCKSVDEar.htm.

III. Block-Wise Statistics-Based Features

In this section, the feature extraction method used in our system will be introduced in detail, which serves as a critical component in our 3D ear identification system.

When SRC or LC-KSVD is adopted as a classification framework, the feature vector extracted from the test sample needs to be sparsely coded over the dictionary whose columns are learned from feature vectors of gallery samples. In the field of face recognition, feature vectors are typically vectorized from raw image pixels and impressive results are obtained [43], [59]. However, these methods actually implicitly require that the test image and the training set must be well aligned. As reported in [44], if the test image has even a small amount of registration error against training images (which is also true for the 3D ear classification problem), the representation coefficients will no longer be informative. To deal with this problem, several studies have been conducted recently. In [44], Wagner *et al.* solve this challenging issue by a series of linear programs that iteratively minimize the sparsity of the registration error. In [62], Peng *et al.* formulate the batch image alignment as searching for a set of transformations that can minimize the rank of the transformed images, which are viewed as columns of a matrix. If Wagner *et al.*'s method [44] or Peng *et al.*'s method [62] is adopted, the misalignment between the test image and images of each training class needs to be rectified explicitly. Obviously, this strategy is quite time consuming for large-scale identification applications.

Even though the 3D ear extraction method proposed by Zhang *et al.* [30] can align ears to some extent, there are still small alignment errors between ear ROIs. Since explicitly registering the test ear sample to the training samples is extremely time-consuming, we need to find a new feature extraction scheme which is robust to mere misalignments while the extracted feature vectors are still highly discriminative. To meet these requirements, we propose a novel 3D feature extraction scheme based on block-wise statistics, whose details will be presented in the following.

A 3D ear can be considered as a surface with various convex and concave structures. We can classify the points on the ear into different types based on their different geometric characteristics. Such a kind of 3D feature is called as ST [61], which has been proved to be highly discriminative. Assume that a 3D ear ROI is represented by $S(x, y, f(x, y))$. Mean curvature H and Gaussian curvature K can be computed as [63]

$$H = \frac{\left(1+f_x^2\right) f_{yy} + \left(1+f_y^2\right) f_{xx} - 2 f_x f_y f_{xy}}{2\left(1+f_x^2+f_y^2\right)^{3/2}} \tag{1}$$

$$K = \frac{f_{xx} f_{yy} - f_{xy}^2}{\left(1+f_x^2+f_y^2\right)^2} \tag{2}$$

TABLE I
ST LABELS DEFINED BY SIGNS OF SURFACE CURVATURES [61]

	$K > 0$	$K = 0$	$K < 0$
$H < 0$	Peak (ST = 1)	Ridge (ST = 2)	Saddle Ridge (ST = 3)
$H = 0$	None (ST = 4)	Flat (ST = 5)	Minimal Surface (ST = 6)
$H > 0$	Pit (ST = 7)	Valley (ST = 8)	Saddle Valley (ST = 9)

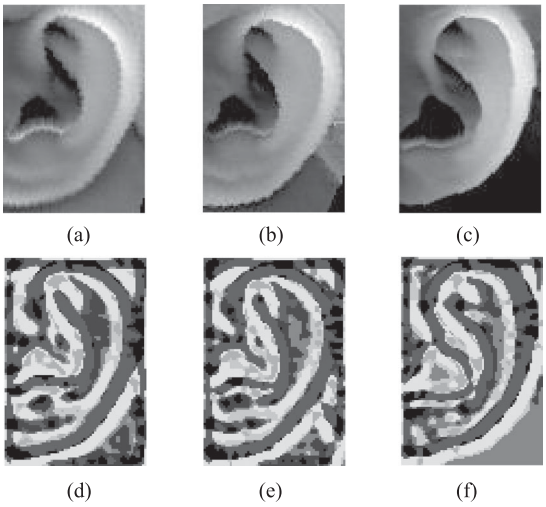


Fig. 1. First row displays three 3D ear ROIs, shown in image format while the second row displays their corresponding ST maps. (a) and (b) are captured from the same ear but in different sessions. (b) and (c) are from different ears.

where $f_x(f_y)$, f_{xx} (f_{yy}, f_{xy}) are the first order and second order partial derivatives, respectively. Details for computing these partial derivatives for range images can be found in Appendix. There are eight fundamental viewpoint independent STs that can be characterized using only the sign of the mean curvature (H) and Gaussian curvature (K) [61]. For completeness, we list their definitions in Table I. In total, 9 STs can be defined, including eight fundamental STs and one special case for $H = 0$ and $K > 0$.

With the above-mentioned procedures, each point in the 3D ear ROI can be classified into one of the 9 STs. Thus, for each 3D ear ROI, we could obtain an ST map, each field of which is an integer from 1 to 9. Examples of ST maps are shown in Fig. 1. In Fig. 1, the first row displays three 3D ear ROIs, shown in image format while the second row displays their corresponding ST maps. Fig. 1(a) and (b) is captured from the same ear but in different sessions while Fig. 1(b) and (c) is captured from different ears.

As a 3D feature, surface type maps are highly discriminative but they are sensitive to small amount of registration errors between the test image and training images. On the other hand, global statistics based features, such as histograms and moment invariants [64], are robust to misalignments but they are not quite discriminative. In order to integrate the merits of these two kinds of feature extraction schemes, we propose to use block-wise ST statistics based features.

Suppose that for a 3D ear ROI, we have computed from it an ST map $\mathcal{M}$. Then, we uniformly divide $\mathcal{M}$ into a set of $p \times p$ blocks. For each block *i*, we compute from it a histogram of STs, denoted by $\mathbf{h}_i$. Obviously, the dimension of $\mathbf{h}_i$ is 9 since there are totally 9 possible STs (see Table I). Finally, all $\mathbf{h}_i$s are concatenated together as a large histogram $\mathbf{h}$, which is considered to be the feature vector. Experimental results have corroborated the efficacy of such a feature extraction scheme (see Section V). The advantages of the proposed feature extraction method are summarized below:

1) *discriminative:* by computing ST maps, it enables the proposed method to be highly discriminative to characterize the rich structures of 3D ears;
2) *robust to mere misalignment:* with concatenation of local statistics of STs, the extracted feature is robust to the mere misalignment existing in 3D ear ROIs; and
3) *low computational cost:* as there are only 9 possible STs, it's rather fast to obtain a local histogram of ST.

IV. LC-KSVD-BASED 3D EAR CLASSIFICATION

By using the proposed feature extraction scheme as presented in Section III, any given 3D ear range image can be represented by a feature vector. With respect to the classification framework, we propose to adopt LC-KSVD [60], whose efficacy and efficiency have been demonstrated in various fields. With LC-KSVD, the supervised information (i.e. class labels) of input signals can be utilized to learn a reconstructive and discriminative dictionary. Each dictionary item will be chosen so that it represents a subset of the training signals ideally from a single class. Thus, each dictionary item can be associated with a particular class label and there is an explicit correspondence between dictionary items and labels. Meanwhile, a multiclass linear classifier can be learned simultaneously.

Given a gallery set comprising of 3D ears, we can compute a feature vector for each sample and then we can define a data matrix $\mathbf{Y}$ as the concatenation of all the extracted feature vectors

$$\mathbf{Y} = [\mathbf{y}_{1,1}, \mathbf{y}_{1,2}, \ldots, \mathbf{y}_{k,n_k}] \in \mathcal{R}^{n \times N} \tag{3}$$

where n is the dimension of the feature vector, k is the number of classes, n_k is the number of samples for class k, and $N = \sum_{j=1}^{k} n_j$ is the total number of samples in the gallery set. The LC-KSVD learning model can be expressed as

$$\begin{aligned}\langle \widehat{\mathbf{D}}, \widehat{\mathbf{W}}, \widehat{\mathbf{A}}, \widehat{\mathbf{X}} \rangle = & \underset{\mathbf{D},\mathbf{W},\mathbf{A},\mathbf{X}}{\arg\min} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \\ & + \alpha \|\mathbf{Q} - \mathbf{A}\mathbf{X}\|_F^2 \\ & + \beta \|\mathbf{H} - \mathbf{W}\mathbf{X}\|_F^2, s.t. \|\mathbf{x}_i\|_0 \le T \end{aligned} \tag{4}$$

where $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_K] \in \mathcal{R}^{n \times K}$ is the learned dictionary, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathcal{R}^{K \times N}$ are the sparse codes of input signals $\mathbf{Y}$, T is the sparsity constraint factor, $\|\mathbf{x}_i\|_0$ counts the non-zero elements in vector $\mathbf{x}_i$, and $\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2$ denotes the reconstruction error. $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_N] \in \mathcal{R}^{K \times N}$ are

the discriminative sparse codes of $\mathbf{Y}$. $\mathbf{q}_i = [\mathbf{q}_i^1, \ldots, \mathbf{q}_i^K]^T = [0, \ldots 1, 1, \ldots 0]^T \in \mathcal{R}^K$ is a discriminative sparse code corresponding to an input signal $\mathbf{y}_i$ since the nonzero values of $\mathbf{q}_i$ occur at those indices where the input signal $\mathbf{y}_i$ and the dictionary item $\mathbf{d}_j (j = 1, \ldots, K)$ share the same label. $\mathbf{A}$ is a linear transformation matrix, which transforms the original sparse codes to be the most discriminative in sparse feature space $\mathcal{R}^K$. Thus, the term $\|\mathbf{Q} - \mathbf{AX}\|_F^2$ represents the discriminative sparse code error, which enforces that the transformed sparse codes $\mathbf{AX}$ approximate the discriminative sparse codes $\mathbf{Q}$. $\mathbf{W}$ denotes the classifier parameters. $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N] \in \mathcal{R}^{k \times N}$ are the class labels of input signals $\mathbf{Y}$. $\mathbf{h}_i = [0, 0, \ldots, 1, ..0, 0]^T \in \mathcal{R}^k$ is a label vector associated to the input signal $\mathbf{y}_i$, where the nonzero position indicates the class label of $\mathbf{y}_i$. Obviously, the term $\|\mathbf{H} - \mathbf{WX}\|_F^2$ represents the classification error. The dictionary $\widehat{\mathbf{D}}$ learned in this way is adaptive to the underlying structure of the training data and can generate discriminative sparse codes $\mathbf{X}$, which can be utilized directly by a linear classifier. The discriminative property of sparse code is very important for the performance of a linear classifier.

For the purpose of optimization, (4) can be rewritten as

$$\langle \widehat{\mathbf{D}}, \widehat{\mathbf{W}}, \widehat{\mathbf{A}}, \widehat{\mathbf{X}} \rangle = \underset{\mathbf{D}, \mathbf{W}, \mathbf{A}, \mathbf{X}}{\arg\min} \left\| \begin{pmatrix} \mathbf{Y} \\ \sqrt{\alpha}\mathbf{Q} \\ \sqrt{\beta}\mathbf{H} \end{pmatrix} - \begin{pmatrix} \mathbf{D} \\ \sqrt{\alpha}\mathbf{A} \\ \sqrt{\beta}\mathbf{W} \end{pmatrix} \mathbf{X} \right\|_F^2$$
$$\text{s.t. } \forall i, \|\mathbf{x}_i\|_0 \leq T\,. \tag{5}$$

Let $\mathbf{Y}_{new} = \left(\mathbf{Y}^T, \sqrt{\alpha}\mathbf{Q}^T, \sqrt{\beta}\mathbf{H}^T\right)^T, \mathbf{D}_{new} = (\mathbf{D}^T, \sqrt{\alpha}\mathbf{A}^T, \sqrt{\beta}\mathbf{W}^T)^T$. The optimization of (5) is equivalent to solving the following problem:

$$\langle \widehat{\mathbf{D}}_{new}, \widehat{\mathbf{X}} \rangle = \underset{\mathbf{D}_{new}, \mathbf{X}}{\arg\min} \|\mathbf{Y}_{new} - \mathbf{D}_{new}\mathbf{X}\|_F^2$$
$$\text{s.t. } \forall i, \|\mathbf{x}_i\|_0 \leq T \tag{6}$$

which can be efficiently solved by the K-SVD algorithm [40]. To solve (6) by using K-SVD, $\mathbf{D}$, $\mathbf{A}$ and $\mathbf{W}$ need to be initialized as $\mathbf{D}^{(0)}$, $\mathbf{A}^{(0)}$, and $\mathbf{W}^{(0)}$ and to this end we use the method proposed in [60]. Specifically, for $\mathbf{D}^{(0)}$, we run several iterations of K-SVD within each class and combine all the outputs of each K-SVD as $\mathbf{D}^{(0)}$. The label of each dictionary item $\mathbf{d}_j$ is initialized based on the class it corresponds to and will remain fixed during the entire dictionary learning process. Sparse codes $\mathbf{X}^{(0)}$ for $\mathbf{Y}$ can be computed by solving

$$\mathbf{x}_i^{(0)} = \underset{\mathbf{x}_i}{\arg\min} \left\| \mathbf{y}_i - \mathbf{D}^{(0)}\mathbf{x}_i \right\|_2^2, s.t.\ \|\mathbf{x}_i\| \leq T \tag{7}$$

where $\mathbf{y}_i$ is the ith column of $\mathbf{Y}$ and $\mathbf{x}_i^{(0)}$ is the ith column of $\mathbf{X}^{(0)}$. For solving (7), we resort to the orthogonal matching pursuit algorithm [65]. To initialize $\mathbf{A}^{(0)}$, the multivariate ridge regression model [66] with the quadratic loss and l_2-norm regularization is employed, which is expressed as

$$\mathbf{A}^{(0)} = \underset{\mathbf{A}}{\arg\min} \left\| \mathbf{Q} - \mathbf{AX}^{(0)} \right\|_F^2 + \lambda_1 \|\mathbf{A}\|_F^2\,. \tag{8}$$

It has a closed-form solution as

$$\mathbf{A}^{(0)} = \mathbf{Q}\left(\mathbf{X}^{(0)}\right)^T \left(\mathbf{X}^{(0)}\left(\mathbf{X}^{(0)}\right)^T + \lambda_1 \mathbf{I}\right)^{-1}. \tag{9}$$

Similarly, for initializing $\mathbf{W}^{(0)}$, we also use the ridge regression model and $\mathbf{W}^{(0)}$ can be computed as

$$\mathbf{W}^{(0)} = \mathbf{H}\left(\mathbf{X}^{(0)}\right)^T \left(\mathbf{X}^{(0)}\left(\mathbf{X}^{(0)}\right)^T + \lambda_2 \mathbf{I}\right)^{-1}. \tag{10}$$

After we get $\widehat{\mathbf{D}}_{new}$ by solving (6), we can obtain $\widehat{\mathbf{D}} = [\mathbf{d}_1, \ldots, \mathbf{d}_K]$ and $\widehat{\mathbf{W}} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$ from $\widehat{\mathbf{D}}_{new}$. However, we cannot directly use $\widehat{\mathbf{D}}$ and $\widehat{\mathbf{W}}$ for testing since $\widehat{\mathbf{D}}$, $\widehat{\mathbf{A}}$, and $\widehat{\mathbf{W}}$ are l_2-normalized in $\widehat{\mathbf{D}}_{new}$ jointly in the LC-KSVD algorithm, i.e., $\forall k, \left\|\mathbf{d}_k^T, \sqrt{\alpha}\mathbf{a}_k^T, \sqrt{\beta}\mathbf{w}_k^T\right\|_2 = 1$. The desired dictionary $\widehat{\mathbf{D}}^*$ and classifier parameters $\widehat{\mathbf{W}}^*$ can be computed as follows:

$$\widehat{\mathbf{D}}^* = \left[\frac{\mathbf{d}_1}{\|\mathbf{d}_1\|_2}, \ldots, \frac{\mathbf{d}_K}{\|\mathbf{d}_K\|_2}\right]$$
$$\widehat{\mathbf{W}}^* = \left[\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|_2}, \ldots, \frac{\mathbf{w}_K}{\|\mathbf{w}_K\|_2}\right]. \tag{11}$$

At the testing stage, given a probe 3D ear scan, we at first detect the ear region and compute from it a feature vector $\mathbf{y}$. Then, we compute its sparse representation $\widehat{\mathbf{x}}$ over the learned dictionary $\widehat{\mathbf{D}}^*$ by solving the following problem:

$$\widehat{\mathbf{x}} = \underset{\mathbf{x}}{\arg\min} \left\| \mathbf{y} - \widehat{\mathbf{D}}^*\mathbf{x} \right\|_2^2, s.t.\ \|\mathbf{x}\|_0 \leq T\,. \tag{12}$$

After $\widehat{\mathbf{x}}$ is obtained, we simply use the linear predictive classifier to estimate a label vector $\mathbf{c} = \widehat{\mathbf{W}}^*\widehat{\mathbf{x}}$. Finally, the label of $\mathbf{y}$ is assigned as the index corresponding to the largest element of $\mathbf{c}$. Our proposed 3D ear identification algorithm is summarized in Table II. Its overall flowchart is illustrated in Fig. 2.

V. Experiments

A. Database and Experimental Protocol

In experiments, we used the UND Collection J2 dataset [67]. This dataset contains 2346 3D side face scans captured from 415 different persons, making it the largest 3D ear scan dataset so far. Those range images were collected using a Minolta Vivid 910 range scanner in high resolution mode. There are variations in pose between them and some images are occluded with hair or ear rings. Each scan is a 640×480 range image. Several scan samples are shown in Fig. 3.

To evaluate the performance of our method, however, we cannot simply conduct experiments on the whole dataset since some classes in UND-J2 have only 2 samples. As pointed out in [43], classification schemes based on sparse coding need sufficient samples for each class in the gallery. Consequently, we virtually created four subsets from UND-J2 for experiments. Specifically, we required that each class should have more than 6, 8, 10, and 12 samples, respectively. For subset 1, we randomly selected from each class 6 samples to form the gallery set and the rest samples were used to form the test set. For subset 2, we randomly selected from each class 8 samples to form the gallery

TABLE II
PROPOSED ALGORITHM FOR 3D EAR IDENTIFICATION

Training phase

Input: A gallery set containing 3D ear ROIs.

Output: Dictionary $\widehat{\mathbf{D}}^*$ and the classifier parameters $\widehat{\mathbf{W}}^*$.

1. For the sample i in the gallery set,

   Extract from it a feature vector $\mathbf{y}_i$;

   Normalize $\mathbf{y}_i$ to have unit l_2-norm;

2. Concatenate all $\{\mathbf{y}_i\}$ as $\mathbf{Y}$;

3. Solve

$$< \widehat{\mathbf{D}}, \widehat{\mathbf{W}}, \widehat{\mathbf{A}}, \widehat{\mathbf{X}} >= \underset{\mathbf{D},\mathbf{W},\mathbf{A},\mathbf{X}}{\arg\min} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \alpha \|\mathbf{Q} - \mathbf{A}\mathbf{X}\|_F^2 + \beta \|\mathbf{H} - \mathbf{W}\mathbf{X}\|_F^2, s.t.\ \|\mathbf{x}_i\|_0 \le T$$

4. Derive $\widehat{\mathbf{D}}^*$ and $\widehat{\mathbf{W}}^*$ based on $\hat{\mathbf{D}}$ and $\hat{\mathbf{W}}$.

Testing phase

Input: A query 3D ear sample, $\widehat{\mathbf{D}}^*$ and $\widehat{\mathbf{W}}^*$.

Output: Identity of the query sample.

1. Extract the ROI from the query sample;
2. Extract the feature vector $\mathbf{y}$ from the ROI;
3. Code $\mathbf{y}$ over $\widehat{\mathbf{D}}^*$ as

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\arg\min} \left\| \mathbf{y} - \widehat{\mathbf{D}}^* \mathbf{x} \right\|_2^2, s.t.\ \|\mathbf{x}\|_0 \le T$$

4. Compute the label vector $\mathbf{c} = \widehat{\mathbf{W}}^* \hat{\mathbf{x}}$;
5. Identity($\mathbf{y}$) = the index of the largest element in $\mathbf{c}$.

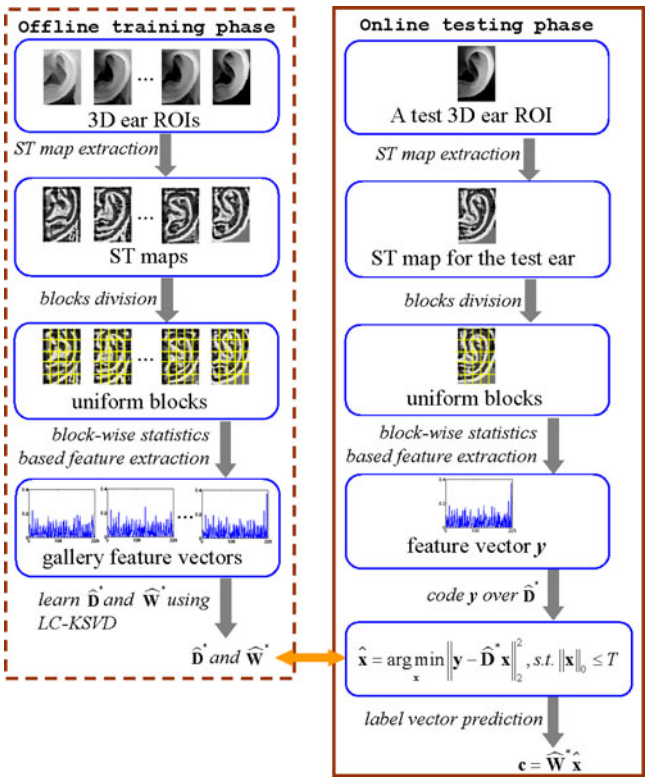


Fig. 2. Illustration for the proposed 3D ear identification approach based on LC-KSVD and block-wise statistics-based features.

set and the rest samples were used to form the test set. For subset 3 and subset 4, similar strategies were used to generate the gallery and test sets. To make it clear, major information about the four subsets used for evaluation is summarized in Table III.

We use the rank-1 recognition rate (R-1 RR) as the performance measure. In addition, the running speed of each competing method was also evaluated. Experiments were performed on a standard HP Z620 workstation with a 3.2 GHZ Intel Xeon E5-1650 CPU and an 8G RAM. The software platform was MATLAB R2013b.

B. Effectiveness of ST Histograms-Based Features

In our proposed 3D ear identification framework, each 3D range image is represented as a feature vector and for feature extraction we propose to use local histograms of STs (LHST) as features. That is, for each range block, we extract from it a histogram of STs and then we concatenate the histograms of all blocks together as the feature vector. In this section, to demonstrate the effectiveness of the proposed feature extraction scheme LHST, we compared its performance with several other feature extraction methods existing in the literature. In order to evaluate the performance of different features, we need to fix the classification approach. In this experiment, with respect to the classification framework, we used the LC-KSVD framework.

The local histogram of STs can be viewed as a kind of local statistics based features. Actually, in the literature there are also other local statistics based features. For example, local binary pattern (LBP) has been testified to be a powerful descriptor for many image classification tasks [68]. When using LBP, actually we regard the range image data as standard image data. In this experiment, when extracting LBP-based features, for each 3D ear ROI, we divided it into uniform blocks, extracted local histogram of LBP from each block, and then concatenated all the histograms to form the final feature vector. An LBP operator can be represented as $LBP_{P,R}^{riu2}$, where "*riu*2" means the use of *r*otation *i*nvariant *u*niform patterns that have transitions at most 2, R is the sampling radius and P is the number of sampling points. We tested three LBP operators $LBP_{8,1}^{riu2}$, $LBP_{16,3}^{riu2}$, and $LBP_{24,5}^{riu2}$, and also their combinations denoted by LBP_m.

Local orientation coding based methods have been verified to be quite successful in the fields of 2D biometrics. For example, CompCode [9], which encodes the local orientation using a set of Gabor filters, is a quite powerful method for matching 2D palmprints. In this experiment, we tested its performance for 3D ear classification. Specifically, for each block, we extract from it a histogram of CompCode and then form the feature vector by concatenating all the local histograms.

In addition, we also evaluated the performance of a local PCA-based feature designed for range images [30]. In such a method, for each point p_i on the range image, neighborhood points will be located at first, based on which PCA is performed. Then, the feature value for p_i is the difference between lengths of the first two principal axes. Finally, the feature map is vectorized as a vector.

Moreover, MCI, GCI and ST maps [10] were computed and then vectorized for performance evaluation as well. Although

Fig. 3. Samples of 3D side face scans in UND-J2 dataset.

TABLE III
SUBSETS USED IN OUR EXPERIMENT

Subset index	♯Classes	Gallery size	Probe size	Total samples
1	127	762	715	1477
2	85	680	461	1141
3	62	620	291	911
4	39	468	168	636

TABLE IV
R-1 RRS BY USING DIFFERENT FEATURES (%)

	Subset 1	Subset 2	Subset 3	Subset 4
$LBP^{riu2}_{8,1}$	74.83	80.69	86.94	92.26
$LBP^{riu2}_{16,3}$	84.62	91.97	95.88	97.02
$LBP^{riu2}_{24,5}$	90.21	94.14	96.56	97.62
LBP_m	88.95	93.93	97.25	97.62
CompCode	90.19	94.66	96.25	98.40
PCA	87.83	91.76	96.22	96.43
MCI	66.99	84.38	85.57	92.26
GCI	66.99	76.79	80.41	88.69
ST	89.23	93.71	96.22	98.21
LHST	**92.86**	**95.88**	**98.63**	**100**

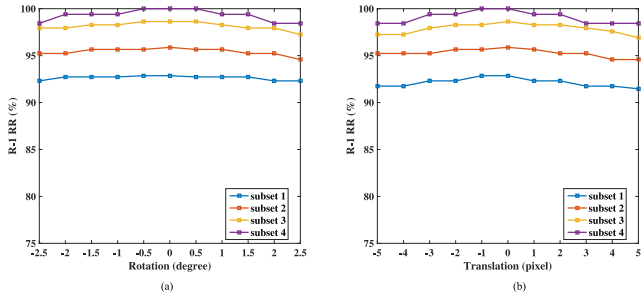


Fig. 4. Robustness of the proposed feature extraction scheme against various degrees of misalignment resulted from (a) rotation and (b) translation.

TABLE V
R-1 RRS BY USING DIFFERENT METHODS (%)

	Subset 1	Subset 2	Subset 3	Subset 4
ICP	83.22	90.02	94.09	95.83
Zhang *et al.* [30]	83.78	90.67	94.50	96.43
SRC_LHST	92.17	94.36	96.56	98.81
LCKSVD_LHST	**92.86**	**95.88**	**98.63**	**100**

these feature extraction schemes are also based on surface curvatures, they fail to cope with the mere misalignment existing in 3D ear ROIs. Details for computing MCI, GCI and ST maps can be found in [10].

The evaluation results are summarized in Table IV. From Table IV, it can be observed that ST outperforms all the other vectorized feature maps; meanwhile, the proposed scheme LHST based on block-wise ST histograms works much better than the rest local statistics based ones. These two points strongly underpin the adoption of ST in our method. Besides, LHST performs better than ST because it further utilizes local histograms for handling misalignment. It indicates that the proposed feature extraction approach is quite qualified in characterizing local shape structures of 3D range data.

To explore the robustness of the proposed feature extraction scheme against misalignment, we considered rotation and translation as the main causes in the task of 3D ear recognition. In this experiment, we simulated different degrees of misalignment using the entire probe 3D ear ROIs from the four subsets. Specifically, we rotated each test sample by -2.5° to 2.5° stepped by 0.5°. For the impact of translation, on the other hand, we translated each one by -5 pixels to 5 pixels at an interval of 1 pixel on both axes. The results are shown in Fig. 4. From Fig. 4, we can see that the proposed LHST is quite robust to the misalignment resulted from various degrees of rotation and translation.

C. Performance Evaluation and Discussions

In this experiment, the performance of several competing methods was evaluated. Our proposed method based on the LC-KSVD classification framework and LHST features is denoted by LCKSVD_LHST. In order to demonstrate the superiority of LC-KSVD as a classification framework, we also tested the performance of the approach which classifies LHST features by using the SRC model [43]. This method is denoted by SRC_LHST. For solving the l_1-minimization problem involved in SRC_LHST, we used the algorithm DALM [69]. Some other state-of-the-art methods for 3D ear matching were also evaluated. They include ICP and Zhang *et al.*'s method [30].

The evaluation results are presented in Tables V and VI. In Table V, we list the R-1 RR achieved by each method on each subset and in Table VI we list the time cost consumed by one identification operation by each method on each subset. Given a test sample, the time cost for one identification operation includes the time consumed by the feature extraction and the time consumed by matching the test feature with the gallery feature set.

TABLE VI
TIME COST FOR ONE IDENTIFICATION OPERATION (SECONDS)

	Subset 1	Subset 2	Subset 3	Subset 4
ICP	5.356×10^5	3.763×10^5	1.876×10^5	1.287×10^5
Zhang *et al.* [30]	2.425	2.424	2.423	2.420
SRC_LHST	0.074	0.070	0.066	0.056
LCKSVD_LHST	**0.058**	**0.034**	**0.033**	**0.018**

Based on the results listed in Tables V and VI, we could have the following findings. At first, with respect to the classification accuracy, the proposed method LCKSVD_LHST performs the best on almost all the subsets. This not only attributes to the robustness of the proposed feature extraction scheme against mere misalignment, but also owes to the discriminative dictionary and the linear classifier jointly learned with LC-KSVD. Particularly, on subset 4 by using our test protocol, the recognition rate of LCKSVD_LHST is 100%, which is quite amazing.

Secondly, SRC_LHST performs much better than Zhang *et al.*'s method [30] though they both exploit the SRC framework for classification. The major difference between SRC_LHST and Zhang *et al.*'s method [30] is that they resort to different schemes for feature extraction. Thus, we can conclude that as a feature extraction scheme, the proposed method LHST is superior to the PCA-based one used in [30].

Thirdly, LCKSVD_LHST performs better than SRC_LHST. The only difference between these two methods is the classification schemes they use; LCKSVD_LHST uses LC-KSVD while SRC_LHST adopts SRC. Hence, the result indicates that as a classification scheme, when the features are extracted by LHST, LC-KSVD performs better than SRC for the task of 3D ear classification.

In addition, in terms of the running speed at the test stage, the proposed method LCKSVD_LHST runs faster than all the other methods evaluated. The computational burden of ICP is extremely heavy, making it not suitable for large-scale identification applications. SRC_LHST runs faster than Zhang *et al.*'s method [30], though they use the same classification criterion. The main reason is that the feature extraction method used in SRC_LHST (i.e., LHST) is much more efficient than the one adopted in [30] (i.e., local PCA-based method). In SRC_LHST, given a test sample, its sparse representation vector is computed at first and then its label is determined by checking the reconstruction residues associated with classes. By contrast, in LCKSVD_LHST, when the sparse representation vector is ready, the label vector can be simply estimated by using a linear predictive classifier, which is much more efficient. That's why the proposed method LCKSVD_LHST is faster than SRC_LHST.

D. Comparison With Other Methods

Besides the methods mentioned in Section V-C, there are also some other state-of-the-art or representative methods in the field of 3D ear recognition, such as Chen and Bhanu [34], Yan and Bowyer [31], Islam *et al.* [32], and Islam *et al.* [33]. However, the source codes of these methods are not publicly available and thus it is nearly impossible for us to accurately re-implement them. Hence, we simply quote the R-1 RRs results originally reported in these papers and summarize them in Table VII. Based on Table VII, we could make some qualitative analysis.

TABLE VII
PERFORMANCE COMPARISON WITH THE OTHER STATE-OF-THE-ART METHODS

Method	Database	Images used (gallery, probe)	Peformance (%)
Chen and Bhanu [34]	UND-F [70]	604 (302, 302)	96.36
Yan and Bowyer [31]	UND-J2	1801 (415, 1386)	97.80
Islam *et al.* [32]	UND-F [70]	200 (100,100)	90.00
Islam *et al.* [33]	UND-J2	830 (415, 415)	93.50
LCKSVD_LHST	UND-J2	1141 (680, 461)	95.58
LCKSVD_LHST	UND-J2	911 (620, 291)	98.63
LCKSVD_LHST	UND-J2	636 (468, 168)	100

At first, based on the published results, it can be seen that Yan and Bowyer's method [31] and Chen and Bhanu's method [34] are the state-of-the-art ones. Yan and Bowyer's method [31] can achieve an R-1 RR 97.8% on a dataset comprising 1801 samples. Secondly, the proposed approach LCKSVD_LHST can achieve quite competitive recognition accuracy with the state-of-the-art ones.

Actually, compared with the state-of-the-art methods [31]–[34], the proposed method LCKSVD_LHST has several inherent advantages. At first, LCKSVD_LHST depends only on 3D range data while the other ones require both the 3D data and the co-registered 2D data. Thus, LCKSVD_LHST is conceptually much simpler and can be used in the case where co-registered 2D data is not available. Secondly, from Table VI it can be seen that LCKSVD_LHST is quite efficient for large-scale identification applications. By contrast, all the other methods evaluated here adopt ICP (or its variants) for matching. When using these methods for identification, it would be necessary to match the test sample to all the gallery samples one by one by performing pair-wise ICP (or its variants). Obviously, they are not computationally efficient, especially when the gallery size is extremely large.

Based on the above discussions, we recommend using the proposed LCKSVD_LHST method for 3D ear identification since such an approach could achieve a distinguished high recognition accuracy while maintaining an extremely low computational complexity. LCKSVD_LHST is quite suitable for large-scale identification applications.

VI. CONCLUSION

In this paper, we proposed a novel method for 3D ear identification, namely LCKSVD_LHST. Our contributions are mainly from two aspects. At first, we are the first to adapt LC-KSVD, a state-of-the-art model for supervised dictionary learning, to the application of 3D ear recognition. Secondly, for feature extraction, we proposed an approach based on local histograms of STs, which is quite effective and robust to small alignment errors.

Experiments conducted on benchmark dataset demonstrate that LCKSVD_LHST could achieve much higher recognition rate than the other competitors evaluated. In addition, its computational complexity is extremely low at the test stage, making it quite suitable for large-scale identification applications.

APPENDIX
IMPLEMENTATION DETAILS

Some details in implementation are presented here. At first, for computing curvatures for range images [see (1) and (2)], partial derivatives with various orders need to be estimated. To reliably estimate partial derivatives, we resort to the scheme proposed in [61]. Specifically, the range image is at first smoothed by using a binomial filter and then partial derivatives are computed by convolving with various predefined window masks. The binomial smoothing filter can be written as $\mathbf{S} = \mathbf{s}\mathbf{s}^T$, where the column vector $\mathbf{s}$ is given by

$$\mathbf{s} = \frac{1}{64}[1 \;\; 6 \;\; 15 \;\; 20 \;\; 15 \;\; 6 \;\; 1]^T. \tag{13}$$

Derivative estimation window masks are defined as $\mathbf{D}_x = \mathbf{d}_0\mathbf{d}_1^T$, $\mathbf{D}_y = \mathbf{d}_1\mathbf{d}_0^T$, $\mathbf{D}_{xx} = \mathbf{d}_0\mathbf{d}_2^T$, $\mathbf{D}_{yy} = \mathbf{d}_2\mathbf{d}_0^T$, and $\mathbf{D}_{xy} = \mathbf{d}_1\mathbf{d}_1^T$, where the column vectors $\mathbf{d}_0$, $\mathbf{d}_1$, and $\mathbf{d}_2$ are given by

$$\begin{aligned}
\mathbf{d}_0 &= \frac{1}{7}[1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1]^T \\
\mathbf{d}_1 &= \frac{1}{28}[-3 \;\; -2 \;\; -1 \;\; 0 \;\; 1 \;\; 2 \;\; 3]^T \\
\mathbf{d}_2 &= \frac{1}{84}[5 \;\; 0 \;\; -3 \;\; -4 \;\; -3 \;\; 0 \;\; 5]^T.
\end{aligned} \tag{14}$$

Then, partial derivative maps of the image $f(x, y)$ are computed as

$$\begin{aligned}
f_x(x,y) &= \mathbf{D}_x * \mathbf{S} * f(x,y) \\
f_y(x,y) &= \mathbf{D}_y * \mathbf{S} * f(x,y) \\
f_{xx}(x,y) &= \mathbf{D}_{xx} * \mathbf{S} * f(x,y) \\
f_{yy}(x,y) &= \mathbf{D}_{yy} * \mathbf{S} * f(x,y) \\
f_{xy}(x,y) &= \mathbf{D}_{xy} * \mathbf{S} * f(x,y)
\end{aligned} \tag{15}$$

where * denotes the convolution operation.

When computing STs, we need to decide whether the mean curvature H (or the Gaussian curvature K) is 0 or not. However, since both H and K take real values, it is quite rare for them to take the value 0 precisely in practice. Thus, in implementation we need to determine a symmetric interval $[-\varepsilon_H, \varepsilon_H]$ (or $[-\varepsilon_K, \varepsilon_K]$) covering 0 for quantization. H (K) is deemed as 0 when its value is covered by the interval $[-\varepsilon_H, \varepsilon_H]$ ($[-\varepsilon_K, \varepsilon_K]$). To make the threshold ε_H (ε_K) be adaptive to different ears, we normalize H (K) by its standard deviation. Such a technical trick was first proposed in [10]. We set $\varepsilon_H = 0.030$ and $\varepsilon_K = 0.015$ in our implementation.

We set $\alpha = 1$, $\beta = 1$ [see (4)], and $p = 10$, respectively.

REFERENCES

[1] A. K. Jain, P. J. Flynn, and A. Ross, *Handbook of Biometrics*. New York, NY, USA: Springer, 2007.

[2] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. New York, NY, USA: Springer, 2003.

[3] F. Liu, D. Zhang, and L. Shen, "Study on novel curvature features for 3D fingerprint recognition," *Neurocomputing*, vol. 168, no. 1, pp. 599–608, Nov. 2015.

[4] M. Yang, P. Zhu, F. Liu, and L. Shen, "Joint representation and pattern learning for robust face recognition," *Neurocomputing*, vol. 168, no. 1, pp. 70–80, Nov. 2015.

[5] X. Shi, Z. Guo, and Z. Lai, "Face recognition by sparse discriminant analysis via joint $L_{2,1}$-norm minimization," *Pattern Recog.*, vol. 47, no. 7, pp. 2447–2453, Jul. 2014.

[6] H. Wechsler, *Reliable Face Recognition Methods-System Design, Implementation and Evaluation*. New York, NY, USA: Springer, 2006.

[7] J. Daugman, "How iris recognition works," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 21–30, Jan. 2004.

[8] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, "Image understanding for iris biometrics: A survey," *Comput. Vis. Image Understanding*, vol. 110, no. 2, pp. 281–307, May 2008.

[9] A. Kong and D. Zhang, "Competitive coding scheme for palmprint verification," in *Proc. IEEE Int. Conf. Pattern Recog.*, Aug. 2004, pp. 520–523.

[10] D. Zhang, G. Lu, W. Li, and N. Luo, "Palmprint recognition using 3-D information," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 39, no. 5, pp. 505–519, Sep. 2009.

[11] D. Zhang, W. Zuo, and F. Yue, "A comparative study of palmprint recognition algorithms," *ACM Comput. Surveys*, vol. 44, no. 1, pp. 2.1–37, Jan. 2012.

[12] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos, "Biometric identification through hand geometry measurements," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1168–1171, Oct. 2000.

[13] M. S. Nixon, T. N. Tan, and R. Chellappa, *Human Identification Based on Gait*. New York, NY, USA: Springer, 2006.

[14] L. Zhang, L. Zhang, D. Zhang, and H. Zhu, "Online finger-knukle-print verification for personal authentication," *Pattern Recog.*, vol. 43, no. 7, pp. 2560–2571, Jul. 2010.

[15] L. Zhang, L. Zhang, D. Zhang, and Z. Guo, "Phase congruency induced local features for finger-knukle-print recognition," *Pattern Recog.*, vol. 45, no. 7, pp. 2522–2531, Jul. 2012.

[16] A. Iannarelli, *Ear Identification*. Paramount, CA, USA: Paramont Publishing Co., 1989.

[17] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1160–1165, Sep. 2003.

[18] D. J. Hurley, M. S. Nixon, and J. N. Carter, "Force field feature extraction for ear biometrics," *Comput. Vis. Image Understanding*, vol. 98, no. 3, pp. 491–512, Jun. 2005.

[19] M. Choras, "Ear biometrics based on geometrical feature extraction," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 5, no. 3, pp. 84–95, Mar. 2005.

[20] J. D. Bustard and M. S. Nixon, "Toward unconstrained ear recognition from two-dimensional images," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 3, pp. 486–494, May 2010.

[21] L. Yuan and Z. Mu, "Ear recognition based on local information fusion," *Pattern Recog. Lett.*, vol. 33, no. 2, pp. 182–190, Jan. 2012.

[22] A. Kumar and T. T. Chan, "Robust ear identification using sparse representation of local texture descriptors," *Pattern Recog.*, vol. 46, no. 1, pp. 73–85, Jan. 2013.

[23] P. Yan and K.W. Bowyer, "Ear biometrics using 2D and 3D images," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog. Workshops*, Jun. 2005, pp. 121–128.

[24] H. Chen and B. Bhanu, "Human ear detection from side face range images," in *Proc. IEEE Int. Conf. Pattern Recog.*, Aug. 2004, pp. 574–577.

[25] H. Chen and B. Bhanu, "Shape model-based 3D ear detection from side face range images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshops*, Jun. 2005, pp. 122–127.

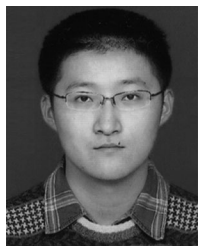
[26] G. Passalis, I. A. Kakadiaris, T. Theoharis, G. Toderici, and T. Papaioannou, "Towards fast 3D ear recognition for real-life biometric applications," in *Proc. IEEE Conf. Adv. Video Signal Based Surveillance*, Sep. 2007, pp. 39–44.

[27] I. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, and T. Theoharis, "Quo vadis: 3D face and ear recognition," in *Face Biometrics Pers. Identif.*, pp. 139–164, 2007.

[28] T. Theoharis, G. Passalis, G. Toderici, and I. A. Kakadiaris, "Unified 3D face and ear recognition using wavelets on geometry images," *Pattern Recog.*, vol. 41, no. 3, pp. 796–804, Mar. 2008.
[29] T. Theoharis, G. Passalis, G. Toderici, and I. A. Kakadiaris, "*A unified approach to 3D face and ear recognition*," presented at the BMVA Symp. Vision-Based Biometrics, London, U.K., Feb. 2007.
[30] L. Zhang, Z. Ding, H. Li, and Y. Shen, "3D ear identification based on sparse representation," *PLoS One*, vol. 9, no. 4, pp. e95506-1–e95506-9, Apr. 2014.
[31] P. Yan and K. W. Bowyer, "Biometric recognition using three-dimensional ear shape," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1297–1308, Aug. 2007.
[32] S. M. S. Islam, R. Davies, A. S. Mian, and M. Bennamoun, "A fast and fully automatic ear recognition approach based on 3D local surface features," in *Proc. 10th Int. Conf. Adv. Concepts Intell. Vis. Syst.*, Oct. 2008, pp. 1081–1092.
[33] S. M. S. Islam, R. Davies, M. Bennamoun, and A. S. Mian, "Efficient detection and recognition of 3D ears," *Int. J. Comput. Vis.*, vol. 95, no. 1, pp. 52–73, Apr. 2011.
[34] H. Chen and B. Bhanu, "Human ear recognition in 3D," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 718–737, Apr. 2007.
[35] P. Yan, K. W. Bowyer, and K. J. Chang, "ICP-based approaches for 3D ear recognition," *Proc. SPIE*, vol. 5779, pp. 282–291, Mar. 2005.
[36] P. Yan and K. W. Bowyer, "Multi-biometrics 2D and 3D ear recognition," in *Proc. Int. Conf. Audio-/Video-Based Biometric Person Authentication*, Jul. 2005, pp. 503–512.
[37] H. Chen and B. Bhanu, "Contour matching for 3D ear recognition," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Jan. 2005, pp. 123–128.
[38] P. J. Besl and N. D. Mckay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
[39] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM Comput. Surveys*, vol. 45, no. 2, pp. 22.1–33, Feb. 2013.
[40] E. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
[41] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, no. 1, pp. 19–60, Jan. 2010.
[42] Z. Zhu, F. Guo, H. Yu, and C. Chen, "Fast single image super-resolution via self-example learning and sparse representation," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2178–2190, Dec. 2014.
[43] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[44] A. Wagner *et al.*, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 2, pp. 372–386, Feb. 2012.
[45] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2009, pp. 1794–1801.
[46] C. Lee, Y. Yang, and H. Chen, "Multipitch estimation of piano music by exemplar-based sparse representation," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 608–618, Jun. 2012.
[47] L. He, D. Tao, X. Li, and X. Gao, "Sparse representation for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 1146–1153.
[48] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, Apr. 2012.
[49] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2008, pp. 1–8.
[50] F. Perronnion, "Universal and adapted vocabularies for generic visual categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1243–1256, Jul. 2008.
[51] I. Remirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2010, pp. 3501–3508.
[52] W. Zhang, A. Surve, X. Fern, and T. Dietterich, "Learning non-redundant codebooks for classifying complex objects," in *Proc. Int. Conf. Mach Learn.*, Jun. 2009, pp. 1241–1248.
[53] N. Zhou, Y. Shen, J. Peng, and J. Fan, "Learning inter-related visual dictionary for object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 3490–3497.
[54] J. Winn, A. Criminisi, and T. Minka, "Object categorization by learned universal visual dictionary," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2005, pp. 1800–1807.
[55] S. Lazebnik and M. Raginsky, "Supervised learning of quantizer codebooks by information loss minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 7, pp. 1294–1309, Jul. 2009.
[56] B. Fulkerson, A. Vedaldi, and S. Soatto, "Localizing objects with smart dictionaries," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2008, pp. 179–192.
[57] N. Slonim and N. Tishby, "Agglomerative information bottleneck," in *Proc. Conf. Neural Inf. Process. Syst.*, Nov. 1999, pp. 617–623.
[58] D. Pham and S. Venkatesh, "Joint learning and dictionary construction for pattern recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 1–8.
[59] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 543–550.
[60] Z. Jiang, Z. Lin, and L. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.
[61] P. J. Besl and R. C. Jain, "Segmentation through variable-order surface fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 2, pp. 167–192, Feb. 1988.
[62] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2233–2246, Nov. 2012.
[63] M. P. Do Carmo, *Differential Geometry of Curves and Surfaces*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1976.
[64] J. Flusser, B. Zitova, and T. Suk, *Moments and Moment Invariants in Pattern Recognition*. Hoboken, NJ, USA: Wiley, 2009.
[65] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2233–2246, Nov. 2012.
[66] G. Golub, P. Hansen, and D. Oleary, "Tikhonov regularization and total least squares," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 1, pp. 185–194, Jan. 1999.
[67] K. W. Bowyer and P. J. Flynn, *ND-Collection J2*. (2007) [Online]. Available: https://sites.google.com/a/nd.edu/public-cvrl/data-sets
[68] T. Ojala, M. Pietikainen, and T. Maenpaa, "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.
[69] J. Yang and Y. Zhang, "Alternating direction algorithms for l_1-problems in compressive sensing," *SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250–278, Jan. 2011.
[70] P. Yan and K. W. Bowyer, "Empirical evaluation of advanced ear biometrics," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshops*, Jun. 2005, pp. 41–48.

Lin Zhang (S'10-M'11-SM'15) received the B.Sc. and M.Sc. degrees in computer science and engineering from Shanghai Jiao Tong University, Shanghai, China, in 2003 and 2006, respectively, and the Ph.D. degree in computing from the Hong Kong Polytechnic University, Hong Kong, China, in 2011.

From March 2011 to August 2011, he was a Research Assistant with the Department of Computing, Hong Kong Polytechnic University. In August 2011, he joined the School of Software Engineering, Tongji University, Shanghai, China, where he is currently an Associate Professor. His current research interests include biometrics, pattern recognition, computer vision, and perceptual image/video quality assessment.



Lida Li received the B.S. degree in software engineering from Tongji University, Shanghai, China, in 2013, where he is currently working toward the M.S. degree.

His research interests include biometrics and machine learning.

Hongyu Li received the B.E. degree from Tongji University, Shanghai, China, in 2000, the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2008, and the Ph.D. degree in computer science from the University of Eastern Finland, Joensuu, Finland, in 2012.

In August 2008, he joined the School of Software Engineering, Tongji University, where he is currently an Associate Professor. His research interests include computer vision, pattern recognition, and motion analysis.

Meng Yang (M'13) received the Ph.D. degree from the Hong Kong Polytechnic University, Hong Kong, China, in 2012.

He was previously a Postdoctoral Fellow with the Computer Vision Laboratory, ETH Zurich, Zurich, Switzerland. He is currently an Associate Professor with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. He has authored or coauthored nine papers for the Association for the Advance of Artificial Intelligence, the Conference on Artificial Intelligence, the IEEE Conference on Computer Vision and Pattern Recognition, the IEEE International Conference on Computer Vision, and the European Conference on Computer Vision, and several journal papers for the *International Journal of Computer Vision*, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and the IEEE TRANSACTIONS ON IMAGE PROCESSING. His research interests include sparse coding, dictionary learning, object recognition, and machine learning.