3D EAR IDENTIFICATION USING LC-KSVD AND LOCAL HISTOGRAMS OF SURFACE TYPES

Lida Li, Lin Zhang, Hongyu Li*

School of Software Engineering
Tongji University
Shanghai 201804, China

ABSTRACT

In this paper, we propose a novel 3D ear classification scheme, making use of the label consistent K-SVD (LC-KSVD) framework. As an effective supervised dictionary learning algorithm, LC-KSVD learns a compact discriminative dictionary for sparse coding and a multi-class linear classifier simultaneously. To use LC-KSVD, one key issue is how to extract feature vectors from 3D ear scans. To this end, we propose a block-wise statistics based scheme. Specifically, we divide a 3D ear ROI into blocks and extract a histogram of surface types from each block; histograms from all blocks are concatenated to form the desired feature vector. Feature vectors extracted in this way are highly discriminative and are robust to mere misalignment. Experimental results demonstrate that the proposed approach can achieve much better recognition accuracy than the other state-of-the-art methods. More importantly, its computational complexity is extremely low at the classification stage.

***Index Terms*—** 3D ear, surface type, local histogram, LC-KSVD, sparse representation

1. INTRODUCTION

Propelled by the requirements of numerous applications, 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 are drawing increasing attention recently because of their high accuracy and robustness in the modern e-world. In the past several decades or so, researchers have exhaustively investigated a number of different biometric identifiers, such as fingerprint, face, iris, palmprint, hand geometry, voice, gait, etc [1].

Among all the members in the big family of biometric identifiers, ear is a new comer but has recently received significant attention due to its non-intrusiveness and ease of data collection. Ear recognition problems can be roughly classified as 2D, 3D, and multimodal 2D plus 3D, according to the types of input data. 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 [2].

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.

2. RELATED WORKS AND OUR CONTRIBUTIONS

2.1. 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 ones are totally based on 3D range data, e.g. [3, 4, 5, 6], while the other ones are based on multi-modal co-registered 2D plus 3D data [7, 8]. The ones based solely on 3D data are reviewed here since they are most relevant to our work. In [3], 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 comprises four steps, step edge detection and thresholding, image dilation, connected-component labeling, and template matching. In their later work [4], Chen and Bhanu proposed a shape-model based technique for locating ears in a side face range image. In [6], 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.

With respect to the 3D ear matching schemes, most of the existing state-of-the-art methods adopt ICP [9] or its variants. 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

*Corresponding author. Email: cslinzhang@tongji.edu.cn

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. Therefore, ICP-based methods are not appropriate for dealing with large-scale identification applications. In [6], Zhang *et al*. tried to solve this problem by making use of 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 $\mathbf{A}$. When a test sample comes, its feature vector $\mathbf{y}$ will be extracted at first and then its identity can be figured out by using SRC which codes $\mathbf{y}$ over the dictionary $\mathbf{A}$.

2.2. Sparse coding and dictionary learning

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 [10], image restoration [11], and object classification [6].

With sparse coding, an input signal $\mathbf{y}$ is approximated by a linear combination of a few items from an overcomplete dictionary $\mathbf{A}$. As pointed out in [12], usually a dictionary learned from the training samples can produce better results than the one using off-the-shelf bases. Several prominent supervised dictionary learning methods have been proposed in the literature. Quite recently, Jiang *et al*. proposed a dictionary learning method, namely label consistent K-SVD (LC-KSVD) [13]. 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.

2.3. 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 ROI (region of interest) extraction, we used the method proposed in [6]. 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 [13], 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 the 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 [14] 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.

3. BLOCK-WISE STATISTICS BASED FEATURES

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 pleasing results could be obtained [15]. However, these methods actually implicitly require that the test image and the training set must be well aligned. As reported in [16], 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 [16], Wagner $et\ al.$ solve this challenging issue by a series of linear programs that iteratively minimize the sparsity of the registration error. In [17], 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 [16] or Peng $et\ al.$'s method [17] 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 $et\ al.$ [6] 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 expect 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 surface type (ST) [14], which has been proved to be highly

Table 1. ST labels defined by signs of surface curvatures [14]

	$K > 0$	$K = 0$	$K < 0$
$H < 0$	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)

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 [18],

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

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

where $f_x(f_y)$, $f_{xx}(f_{yy}, f_{xy})$ are the first order and second order partial derivatives, respectively. There are 8 fundamental viewpoint independent surface types that can be characterized using only the sign of the mean curvature (*H*) and Gaussian curvature (*K*) [14]. For completeness, we list their definitions in Table 1. Totally, 9 STs can be defined, including 8 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 a ST map, each field of which is an integer from 1 to 9. Fig. 1 shows some examples.

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, 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 a 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 surface types, denoted by $\boldsymbol{h}_i$. Finally, all $\boldsymbol{h}_i$s are concatenated together as a large histogram $\boldsymbol{h}$, which is considered to be the feature vector. Experimental results have corroborated the efficacy of such a feature extraction scheme (see Section 5).

4. LC-KSVD BASED 3D EAR CLASSIFICATION

By using the proposed feature extraction scheme as presented in Section 3, any given 3D ear range image can be represented by a feature vector. In terms of the classification framework, we propose to adopt LC-KSVD [13], whose efficacy and efficiency have been demonstrated in several various fields.

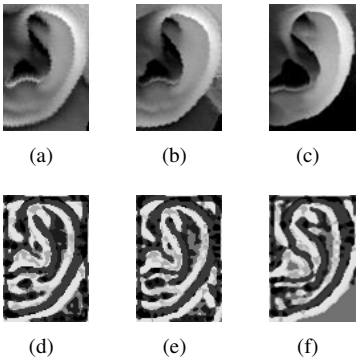

Fig. 1. The first row displays three 3D ear ROIs, shown in image format while the second row displays their corresponding ST maps. (a) and (b) are captured from the same ear but in different sessions. (b) and (c) are from different ears.

Given a gallery set comprising 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 feature vectors,

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

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,

$$\begin{aligned} \left\langle \hat{\mathbf{D}}, \hat{\mathbf{A}}, \hat{\mathbf{W}}, \hat{\mathbf{X}} \right\rangle = \underset{\mathbf{D},\mathbf{A},\mathbf{W},\mathbf{X}}{\operatorname{argmin}} & \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \\ & + \alpha \|\mathbf{Q} - \mathbf{A}\mathbf{X}\|_F^2 \\ & + \beta \|\mathbf{H} - \mathbf{W}\mathbf{X}\|_F^2, s.t.\ \forall i, \|\mathbf{x}_i\|_0 \leq \tau \end{aligned} \quad (4)$$

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 $\mathbf{Y}$, τ is the sparsity constraint factor, $\|\mathbf{x}_i\|_0$ counts the nonzero 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 $\mathbf{Y}$. $\mathbf{q}_i = \left[\mathbf{q}_i^1, \ldots, \mathbf{q}_i^K\right]^T = [0, \ldots, 1, 1, \ldots, 0]^T \in \mathcal{R}^K$ 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 the most discriminative in sparse feature space $\mathcal{R}^K$. Thus, the term $\|\mathbf{Q} - \mathbf{A}\mathbf{X}\|_F^2$ represents the discriminative sparse code error, which enforces that the transformed sparse codes $\mathbf{A}\mathbf{X}$ 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 $\mathbf{Y}$. $\mathbf{h}_i = [0, \ldots, 1, \ldots, 0]^T \in \mathcal{R}^k$ is a label vector associated to the input signal $\mathbf{y}_i$. Obviously, the term $\|\mathbf{H} - \mathbf{W}\mathbf{X}\|_F^2$ represents the classification error. The dictionary $\hat{\mathbf{D}}$ learned in this way is adaptive to the underlying structure of the training

data and can generate discriminative sparse codes, which can be utilized directly by a linear classifier.

For optimization purposes, Eq. 4 can be rewritten as,

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

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

$$\left\langle \hat{\mathbf{D}}_{new}, \hat{\mathbf{X}} \right\rangle = \underset{\mathbf{D}_{new},\mathbf{X}}{\operatorname{argmin}} \|\mathbf{Y}_{new} - \mathbf{D}_{new}\mathbf{X}\|_F^2, \quad s.t.\ \forall i, \|\mathbf{x}_i\|_0 \leq \tau \tag{6}$$

which can be efficiently solved by the K-SVD algorithm [19]. The tricks to initialize $\mathbf{D}$, $\mathbf{A}$ and $\mathbf{W}$ can be found in [13]. After we get $\hat{\mathbf{D}}_{new}$ by solving Eq. 6, we can obtain $\hat{\mathbf{D}} = [\mathbf{d}_1, \ldots, \mathbf{d}_K]$ and $\hat{\mathbf{W}} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$ from $\hat{\mathbf{D}}_{new}$. However, we cannot directly use $\hat{\mathbf{D}}$ and $\hat{\mathbf{W}}$ for testing since $\hat{\mathbf{D}}$, $\hat{\mathbf{A}}$ and $\hat{\mathbf{W}}$ are l_2-normalized in $\hat{\mathbf{D}}_{new}$ 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 $\hat{\mathbf{D}}^*$ and classifier parameters $\hat{\mathbf{W}}^*$ can be computed as follows,

$$\hat{\mathbf{D}}^* = \left[\frac{\mathbf{d}_1}{\|\mathbf{d}_1\|_2}, \cdots, \frac{\mathbf{d}_K}{\|\mathbf{d}_K\|_2}\right], \quad \hat{\mathbf{W}}^* = \left[\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|_2}, \cdots, \frac{\mathbf{w}_K}{\|\mathbf{w}_K\|_2}\right] \tag{7}$$

At the test 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 $\hat{\mathbf{x}}$ over the learned dictionary $\hat{\mathbf{D}}^*$ by solving the following problem,

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\|\mathbf{y} - \hat{\mathbf{D}}^*\mathbf{x}\right\|_2^2, \ s.t.\ \|\mathbf{x}\|_0 \leq \tau \tag{8}$$

For solving Eq. 8, we resort to the orthogonal matching pursuit (OMP) algorithm [20]. After $\hat{\mathbf{x}}$ is obtained, we simply use the linear predictive classifier to estimate a label vector $\mathbf{c} = \hat{\mathbf{W}}^*\hat{\mathbf{x}}$. Finally, the label of $\mathbf{y}$ is assigned as the index corresponding to the largest element of $\mathbf{c}$. The overall flowchart of our 3D ear recognition method is illustrated in Fig. 2.

5. EXPERIMENTS

5.1. Database and experimental protocol

In experiments, we used the UND Collection J2 dataset [21], which contains 2346 3D side face scans captured from 415 persons, making it the largest 3D ear scan dataset so far.

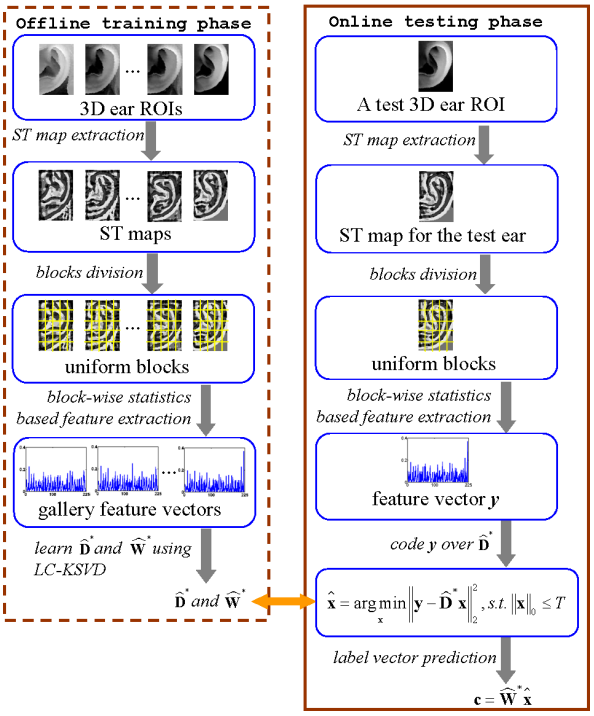

Fig. 2. Illustration for the proposed 3D ear identification approach based on LC-KSVD and block-wise statistics based features.

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 and classification schemes based on sparse coding need sufficient samples for each class in the gallery [25]. 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. Similar strategies were applied to the rest subsets. To make it clear, major information about the four subsets is summarized in Table 2.

We evaluated the recognition rate achieved by each competing method as well as their running speed. Experiments were performed on a standard HP Z620 workstation with a 3.2GHz Intel Xeon E5-1650 CPU and an 8G RAM. The software platform was Matlab R2013b.

5.2. 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. In our implementation, p was set to 10 for a range image. To demonstrate the effectiveness of LHST, we compared its performance with several other kinds of features existing in the literature. In order to evaluate the performance of different features, LC-KSVD framework was adopted as a

Table 2. Subsets used in our experiment

subset idx	♯classes	gallery size	probe size	total
1	127	762	715	1477
2	85	680	461	1141
3	62	620	291	911
4	39	468	168	636

Table 3. Recognition rates 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
LHST	**92.86**	**95.88**	**98.63**	**100**

fixed classification approach.

At first, we compared LHST to LBP (local binary pattern), another local statistics-based feature vector that has been testified to be a powerful descriptor for many image classification tasks [22]. 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 $LBP^{riu2}_{P,R}$, where "*riu2*" means the use of *rotation invariant uniform* 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 $LBP^{riu2}_{8,1}$, $LBP^{riu2}_{16,3}$ and $LBP^{riu2}_{24,5}$, and also their combinations denoted by LBP_m.

Besides, we tested the performance of CompCode [23] for 3D ear classification. In addition, the performance of a local PCA-based feature [6] was also evaluated in our experiment.

The evaluation results are summarized in Table 3. From Table 3, it can be clearly observed that the proposed method LHST based on block-wise ST histograms performs much better than all the other methods evaluated. It indicates that such a feature extraction scheme have a stronger capability in characterizing local shape structures of 3D range data.

5.3. Performance evaluation and discussions

In this experiment, the performance of all the 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 sparse representation-based classification model [15]. This method is denoted by SRC_LHST.

Table 4. Recognition rates by using different identification methods (%)

	subset 1	subset 2	subset 3	subset 4
ICP	83.22	90.02	94.09	95.83
Zhang *et al*. [6]	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**

Table 5. Time cost for one identification operation (second)

	subset 1	subset 2	subset 3	subset 4
ICP($\times 10^5$)	5.356	3.763	1.876	1.287
Zhang *et al*. [6]	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**

We also evaluated some state-of-the-art methods for 3D ear matching, including ICP and Zhang *et al*'.s method [6].

In Table 4, we list the recognition rate achieved by each method on each subset and in Table 5 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 feature matching.

Based on the performance results, we could have the following findings. At first, with respect to the classification accuracy, the proposed method LCKSVD_LHST performs the best on all the subsets. 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 [6] though they both exploit the SRC framework for classification. The major difference between SRC_LHST and Zhang *et al*.'s method [6] is that they resort to different schemes for feature extraction. This fact indicates that as a feature extraction scheme, the proposed method LHST is superior to the PCA-based one used in [6].

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 intimates 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, the proposed method runs greatly faster than all the other methods evaluated at the test stage. 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 [6], though they use the same classification criterion. The main reason is that the feature extraction method used in SRC_LHST is much more efficient than the one adopted in [6]. 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.

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.

6. CONCLUSIONS AND FUTURE WORK

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 the LCKSVD model to the application of 3D ear recognition. Secondly, we proposed an approach based on local histograms of surface types for feature extraction, 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. In future work, we will further analyze LHST by comparing it with some other features, such as HOG features used in [5].

7. ACKNOWLEDGEMENT

This work is supported by the Natural Science Foundation of China under grant no. 61201394 and the Shanghai Pujiang Program under grant nos. 13PJ1408700 and 14PJ1408100.

8. REFERENCES

[1] S.Z. Li, *Encyclopedia of Biometrics*, Springer, 2009.

[2] P. Yan and K.W. Bowyer, "Ear biometrics using 2D and 3D images," in *CVPR Workshops*, 2005, p. 121.

[3] H. Chen and B. Bhanu, "Human ear detection from side face range images," in *ICPR*, 2004, pp. 574–577.

[4] H. Chen and B. Bhanu, "Shape model-based 3D ear detection from side face range images," in *CVPR Workshops*, 2005, p. 122.

[5] J. Zhou, S. Cadavid, and M. Abdel-Mottaleb, "Histograms of categorized shapes for 3D ear detection," in *BTAS*, 2010, pp. 1–6.

[6] L. Zhang, Z. Ding, H. Li, and Y. Shen, "3D ear identification based on sparse representation," *PLoS ONE*, vol. 9, no. e95506, pp. 1–9, 2014.

[7] P. Yan and K.W. Bowyer, "Biometric recognition using three-dimensional ear shape," *IEEE Trans. PAMI*, vol. 29, no. 8, pp. 1297–1308, 2007.

[8] H. Chen and B. Bhanu, "Human ear recognition in 3D," *IEEE Trans. PAMI*, vol. 29, no. 4, pp. 718–737, 2007.

[9] P.J. Besl and N.D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. PAMI*, vol. 14, no. 2, pp. 239–256, 1992.

[10] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. IP*, vol. 15, no. 12, pp. 3736–3745, 2006.

[11] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *CVPR*, 2008, pp. 1–8.

[12] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. PAMI*, vol. 34, no. 4, pp. 791–804, 2012.

[13] Z. Jiang, Z. Lin, and L. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. PAMI*, vol. 35, no. 11, pp. 2651–2664, 2013.

[14] P.J. Besl and R.C. Jain, "Segmentation through variable-order surface fitting," *IEEE Trans. PAMI*, vol. 10, no. 2, pp. 167–192, 1988.

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

[16] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Trans. PAMI*, vol. 34, no. 2, pp. 372–386, 2012.

[17] 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. PAMI*, vol. 34, no. 11, pp. 2233–2246, 2012.

[18] M.D. Carmo, *Differential Geometry of Curves and Surfaces*, Prentice Hall, 1976.

[19] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. SP*, vol. 54, no. 1, pp. 4311–4322, 2006.

[20] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. PAMI*, vol. 34, no. 11, pp. 2233–2246, 2007.

[21] "CVRL Datasets," http://www3.nd.edu/ cvrl.

[22] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. PAMI*, vol. 24, no. 7, pp. 971–987, 2002.

[23] A. Kong and D. Zhang, "Competitive coding scheme for palmprint verification," in *ICPR*, 2004, pp. 520–523.