Multi-dictionary based Collaborative Representation: With Applications to 3D Ear and 3D Palmprint Identification

Anonymous ECCV submission

Paper ID 173

Abstract. With the emergence of reliable and inexpensive 3D scanners, recent years have witnessed a growing interest in developing 3D biometrics systems. As a commonsense, matching algorithms are crucial for such systems. In this paper, we focus on investigating identification methods for two 3D biometric identifiers, 3D ear and 3D palmprint. Specifically, we propose a Multi-Dictionary based Collaborative Representation (MDCR) framework for classification, which can reduce the negative effects aroused by some local regions. With MDCR, a range map is partitioned into overlapping blocks and from each block, a feature vector is extracted. At the dictionary construction stage, feature vectors from blocks having the same locations in gallery samples can form a dictionary and accordingly, multiple dictionaries are obtained. Given a probe sample, by coding its each feature vector on the corresponding dictionary, multiple class labels can be obtained and we use a voting scheme to make the final decision. In addition, a novel patch-wise statistics based feature extraction scheme is proposed, combining the range image’s local surface type information and local dominant orientation information. The effectiveness of the proposed approach has been corroborated by extensive experiments conducted on benchmark datasets. Source codes are available at http://*.

Keywords: Dictionary learning, collaborative representation, 3D ear identification, 3D palmprint identification

1 Introduction

Propelled by the requirements of various applications, such as immigration control, aviation security, or safeguarding of financial transactions, recognizing the identity of a person with high confidence has become the goal of many current research endeavors [1]. To solve such a problem, biometrics based approaches have received considerable attention. In the past decades, researchers have exhaustively investigated a number of various biometric identifiers, such as fingerprint [2], face [3–5], iris [6–8], palmprint [9–13], etc. To date, the majority of implemented biometrics systems are based on 2D images. Despite the great efforts devoted over the past decades, personal authentication using 2D images is still a great challenge. The emergence of reliable and inexpensive 3D scanners has
provided new opportunities for researchers to exploit 3D shape information to perform identity recognition. Compared with 2D image data, 3D data samples have some inherent advantages. For example, they are less sensitive to illumination variations, pose changes, and surface contamination. Moreover, they are relatively more difficult to be copied and counterfeited.

In this paper, we focus on two specific 3D biometric identifiers, 3D ear and 3D palmprint, whose associated recognition systems usually share a common architecture. As shown in Fig. 1, a typical 3D ear or 3D palmprint recognition system comprises the following components, range data acquisition, preprocessing and ROI (region of interest) extraction, feature extraction, and classification. In most cases, the acquired 3D ear or palmprint data is actually a range image. In this work, we suppose that ROI maps for 3D ears or palmprints have already been available and we solely focus on how to devise a universal feature representation and classification scheme, which can be explored to classify 3D ears and 3D palmprints both. For 3D ear and 3D palmprint ROI extraction, readers can refer to [14, 15] and [16], respectively, for details.

The remainder of this paper is organized as follows. Section 2 summarizes related works and our contributions. Section 3 introduces our patch-wise statistics based feature extraction scheme. Section 4 presents the structure of our multi-dictionary based collaborative representation framework. Experimental results are reported in section 5. Finally, section 6 concludes the paper.

2 Related works and our contributions

In this section, we will at first briefly review some representative approaches for matching 3D ears or palmprints. Then, our motivations and contributions will be presented.
2.1 Matching methods for 3D ears and 3D palmprints

In order to detect and segment ear regions from the original profile range images, various different ideas have been proposed [17, 18, 14, 19, 20, 16]. By contrast, with respect to the 3D ear matching schemes, most of the state-of-the-art methods, such as [21, 14, 18, 20], adopts ICP (iterative closest point) [22] or its variants. While ICP is an appealing approach for 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. Quite recently, Zhang et al. [15] tried to solve the 3D ear identification problem using the sparse representation based classification framework (SRC) [3]. In their method, feature vectors are extracted from ear samples in the gallery set and they form an overcomplete dictionary $A$. 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 $A$ from the same class. When a test sample is presented, its feature vector $y$ is extracted at first and then $y$ is coded over the dictionary $A$; the identity of $y$ can be figured out by checking which class leads to the minimum representation error.

For matching 3D palmprints, researchers have devoted a great deal of efforts. In [16], mean curvature images (MCI), Gaussian curvature images (GCI), and surface type (ST) maps were computed and for matching the authors used the normalized Hamming distance. In [23], Li et al. extracted three levels of 2D and 3D palmprint features, including shape features, principal line features, and texture features. To account for small alignment errors, they performed alignment refinement to the feature maps by using ICP. The apparent drawback of this method lies in its high computational complexity. In their another work [24], Li et al. computed the MCI from the original range data at first and then extracted both line and orientation features from MCI. After that, two types of features were fused at either score level or feature level for matching. In [25], Zhang et al. at first extracted surface curvature maps and then used the normalized local correlation for matching. In [26], Yang et al. utilized the shape index representation to describe the geometry of local regions in a 3D palmprint and they extracted LBP (local binary patterns) and Gabor wavelet features from the shape index image. Then, those two features were finally fused at score level. In [27], for feature extraction, Liu and Li applied the OLOF (orthogonal line ordinal feature) operator [11] on MCI derived from a 3D palmprint. To account for small misalignment between two palmprint samples, they tried to use a cross correlation based method to register two feature maps. Quite recently, Zhang et al. [28] pointed out that all the abovementioned methods are only suitable for...
one-to-one verification applications, but not suitable for large-scale one-to-many identification applications. The main reason is that they all adopt a brute-force searching strategy for identification. Moreover, for dealing with the mere misalignment between two ROIs, they used the multi-translation-based matching [16, 24] or explicit registration techniques [23, 27], both of which are not quite computationally efficient. As a solution, Zhang et al. [28] proposed a new method for 3D palmprint identification, namely CR.L2, which makes use of CRC_RLS (collaborative representation based classification with regularized least square) [29] as the classification framework. Additionally, they proposed a patch-wise statistics based feature extraction scheme, which is quite effective and robust to small misalignments.

2.2 Our motivations and contributions

Fig. 2. (a) shows the classification process of CR-based methods while (b) demonstrates the idea of the proposed MDCR classification framework.

Since both SRC [3] and CRC_RLS [29] are based on the collaborative representation (CR, i.e., using gallery samples of all classes to represent the probe sample), we refer them as CR-based classification frameworks. Having investigated the literature, we find that CR-based approaches can get the state-of-the-art
results for 3D ear or palmprint identification [15, 28]. In existing CR-based 3D ear or palmprint classification approaches, a single feature vector is extracted, which is actually a type of holistic representation. And accordingly, a single dictionary is constructed from the gallery set. With such a sample representation scheme, when local deformations, corruptions, or occlusions, different from those existing in gallery samples, happen in the probe sample, its feature vector will be affected and consequently will make the representation coefficients less informative.

To deal with this issue and to better explore the discriminant information embedded in data samples, in this paper, we propose to use multiple feature vectors extracted from local blocks to represent a range image. Accordingly, multiple dictionaries are constructed from the gallery set, each of which is composed of feature vectors extracted from blocks having the same locations in different samples. Given a probe sample, its class label can be determined by solving multiple CR-based classification problems. The proposed classification scheme is termed as Multi-Dictionary based Collaborative Representation, MDCR for short.

We use a real example to demonstrate the design rationale of MDCR, as illustrated in Fig. 2. Fig. 2(a) shows the classification process of the CR-based approaches, while Fig. 2(b) illustrates the idea of the proposed MDCR. Suppose that we have a test 3D ear sample, whose ground-truth class label is \(g\). Denote its feature vector by \(y\). With the CR-based classification scheme, a single dictionary \(D\) is built from gallery samples. \(y\) is coded on \(D\) to get its representation coefficients. Then, \(y\)'s class label is determined by checking which class yields the least reconstruction residual. In this example, the least reconstruction residual happens on class \(h\), and thus the test sample is finally misclassified as class \(h\). With the proposed MDCR classification scheme, the test sample is partitioned into \(N\) blocks \(B_i (i=1 \sim N)\). In this example, we set \(N\) as 4 for simplicity. Accordingly, four feature vectors \(y_i (i=1 \sim 4)\) are extracted from blocks \(B_i (i=1 \sim 4)\). Actually, at the dictionary construction stage, four dictionaries \(D_i (i=1 \sim 4)\) have already been constructed from blocks of gallery samples and the blocks generating \(D_i\) have the same locations as \(B_i\). By coding \(y_i\) on \(D_i\), four class label predictions are obtained, \(k, g, g, \) and \(h\). Finally, by using a simple majority-based voting scheme, the class label of the test sample is correctly determined as \(g\). From this example, it can be seen that using MDCR, multiple label predictions can be obtained for a test sample based on its multiple blocks and the final decision is made using a voting scheme. Such a classification strategy can reduce negative effects brought by “bad” local regions (i.e., regions with local corruptions, occlusions, or deformations) in the test sample.

Another contribution of our work is that we propose a novel patch-wise statistics based feature extraction scheme, which combines the range image’s local surface type information and local dominant orientation information. To encode local dominant orientation, we resort to the competitive coding (CompCode) scheme [10]. The proposed feature extraction scheme is referred as Local Histograms of Surface Types and CompCodes, LH_STCC for short.
The effectiveness and efficiency of the proposed approach have been corroborated by extensive experiments conducted on benchmark datasets. To make the results fully reproducible, Matlab source codes have been made publicly available at http://*.

3 LH_STCC: A novel feature extraction scheme for range images

Using the proposed MDCR classification framework, each range image will be partitioned into a set of blocks and a feature vector needs to be extracted from each block. Since there exists mere misalignment between two ROIs, it is highly desired that the extracted feature vectors are robust to small misalignments while maintaining a high discriminant capability. To meet these requirements, Zhang et al. [28] proposed a patch-wise statistics based feature extraction scheme, which makes use of local surface type information of the range image. In this paper, we extend Zhang et al.’s idea by integrating local surface type information with local dominant orientation information. Details of our feature extraction method are presented as follows.

A range image $R$ can be considered as a surface with various convex and concave structures. Its points can be labeled as different “surface types” according to their intrinsic geometric characteristics. Assume that the range map is represented by $R(x, y, f(x, y))$, where $f(x, y)$ denotes the depth value at $(x, y)$. Its mean curvature $H$ and Gaussian curvature $K$ at $(x, y)$ can be computed as [30],

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

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

where $f_x(f_y)$, $f_{xx}(f_{yy}, f_{xy})$ are the first order and second order partial derivatives, respectively. Nine different surface types (STs) can be defined based on signs of $H$ and $K$, as listed in Table 1. Thus, from $R$, we can obtain a ST map $S$, each field of which is an integer from 1 to 9.

<table>
<thead>
<tr>
<th>$H$</th>
<th>$K$</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt; 0$</td>
<td>$&gt; 0$</td>
<td>Peak (ST=1)</td>
</tr>
<tr>
<td>$= 0$</td>
<td>$= 0$</td>
<td>Ridge (ST=2)</td>
</tr>
<tr>
<td>$&gt; 0$</td>
<td>$&gt; 0$</td>
<td>Saddle Ridge (ST=3)</td>
</tr>
<tr>
<td>$&lt; 0$</td>
<td>$&lt; 0$</td>
<td>None (ST=4)</td>
</tr>
<tr>
<td>$= 0$</td>
<td>$&lt; 0$</td>
<td>Flat (ST=5)</td>
</tr>
<tr>
<td>$&gt; 0$</td>
<td>$&lt; 0$</td>
<td>Minimal Surface (ST=6)</td>
</tr>
<tr>
<td>$&gt; 0$</td>
<td>$&gt;$ 0</td>
<td>Pit (ST=7)</td>
</tr>
<tr>
<td>$&lt;$ 0</td>
<td>$&gt;$ 0</td>
<td>Valley (ST=8)</td>
</tr>
<tr>
<td>$&lt; 0$</td>
<td>$&lt;$ 0</td>
<td>Saddle Valley (ST=9)</td>
</tr>
</tbody>
</table>

On the other hand, the local dominant orientation has been proved to be quite discriminative in the fields of 2D biometrics. Thus, we regard the range
map $\mathbf{R}$ as a 2D gray-scale image and use the Gabor filter based CompCode [10] to extract its local dominant orientation information. 2D Gabor filter is defined as,

$$
G(x, y) = \exp \left( -\frac{1}{2} \left( \frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_y^2} \right) \right) \cdot \exp \left( \frac{i 2\pi x'}{\lambda} \right)
$$

(3)

where $x' = x \cos \theta + y \sin \theta$, $y' = -x \sin \theta + y \cos \theta$. In Eq. (3), $\lambda$ represents the wavelength of the sinusoid factor, $\theta$ represents the orientation of the normal to the parallel stripes of the Gabor function, $\sigma_x$ and $\sigma_y$ are the standard deviations of the 2D Gaussian envelop. CompCode assumes that every image pixel resides on a negative “line” and it extracts the orientation of the line using a set of real Gabor filters with different orientations. Denote by $G_R$ the real part of the Gabor filter $G$. With a series of $G_R$s sharing the same parameters, except the parameter of orientation, local dominant orientation information of $\mathbf{R}$ at the position $(x, y)$ can be extracted and coded. This competitive coding process can be expressed as,

$$
\text{CompCode}(x, y) = \arg \min_j \{ \mathbf{R}(x, y) \star G_R(x, y, \theta_j) \}
$$

(4)

where $\star$ stands for the convolution and $\theta_j = j/J, j = 0 \sim J - 1$. $J$ represents the number of orientations and is usually set as 6. We denote the CompCode map computed from $\mathbf{R}$ by $\mathbf{C}$.

Then, based on $\mathbf{S}$ (the surface type map) and $\mathbf{C}$ (the CompCode map), we resort to the patch-wise statistics based scheme [28] to build the feature vector. Specifically, we uniformly partition $\mathbf{S}$ ($\mathbf{C}$) into a set of regular patches. For each patch $i$, we compute from it a normalized histogram of surface types

**Fig. 3.** Illustration for the proposed feature extraction scheme LH_STCC.
(CompCodes), denoted by $s_i$ ($c_i$). Obviously, the dimension of $s_i$ ($c_i$) is 9 (6) since there are totally 9 (6) surface types (CompCodes). Finally, all the $s_i$s and $c_i$s are concatenated together as a large histogram $h$, which is taken as the feature vector for the range image $R$. The proposed feature extraction algorithm is referred as Local Histograms of Surface Types and CompCodes, LH STCC for short. The flowchart illustrating the building process of LH STCC is shown in Fig. 3.

4 MDCR: Multi-dictionary based collaborative representation

In this section, our proposed MDCR classification framework is presented. Suppose that there is a gallery range image set, comprising $M$ samples from $c$ classes. For each sample $i$, we uniformly partition it into $N$ overlapping blocks $B_{ij}$ ($j = 1 \sim N$). A feature vector $v_{ij}$ can be extracted from each $B_{ij}$ using the feature extraction algorithm LH STCC described in Sect. 3. Then, all the feature vectors $\{v_{ij}\}_{i=1}^M$ are stacked together to form a dictionary $D_j$,

$$D_j = [v_{1,j}, v_{2,j}, \ldots, v_{M,j}] \in \mathbb{R}^{d \times M} \quad (5)$$

where $d$ is the dimension of the feature vector extracted from each block. In this way, we can obtain $N$ dictionaries $\{D_j\}_{j=1}^N$ from the gallery set.

Given a probe range image, we at first partition it into $N$ $p \times p$ blocks $\{T_j\}_{j=1}^N$ as we performed to gallery samples. From each block $T_j$, a feature vector $y_j$ ($j = 1 \sim N$) is extracted using the algorithm LH STCC. Then, $y_j$ can be coded as a linear combination of the column vectors in $D_j$, $y_j$'s representation coefficient can be computed by solving a SRC problem [3],

$$x_j^* = \arg \min_{x_j} \{\|y_j - D_j x_j\|_2^2 + \lambda_1 \|x_j\|_1\} \quad (6)$$

or by solving a CRC RLS problem [29],

$$x_j^* = \arg \min_{x_j} \{\|y_j - D_j x_j\|_2^2 + \lambda_2 \|x_j\|_2^2\} \quad (7)$$

Some previous studies have shown that in many visual classification problems, CRC RLS could achieve comparable recognition accuracy with SRC [29, 28]. On the other hand, CRC RLS is much more efficient than SRC since it has a simple closed-form solution while solving SRC will involve a costly iterative optimization. Thus, in this paper, we adopt Eq. (7) to compute $y_j$'s representation coefficients. It can be easily verified that Eq. (7) has a closed-form solution as:

$$x_j^* = (D_j^T D_j + \lambda_2 I)^{-1} D_j^T y_j \quad (8)$$

where $I \in \mathbb{R}^{M \times M}$ is an identity matrix. Let $P = (D_j^T D_j + \lambda_2 I)^{-1} D_j^T$. Clearly, $P$ is independent of $y_j$ and can be pre-computed solely based on the gallery
set. Then, we can get a class label prediction $c_j$ for the probe range image by checking which class can yield the least reconstruction error for $y_j$,

$$c_j = \arg \min_i \| y_j - D_j \delta_i(x_j^*) \|^2_2$$  \hspace{1cm} (9)

where $\delta_i(x_j^*)$ is a new vector whose only nonzero entries are the entries in $x_j^*$ that are associated with class $i$. $e_{ij} = \| y_j - D_j \delta_i(x_j^*) \|^2_2$ represents the reconstruction error of $y_i$ using gallery samples from class $i$. After processing all the feature vectors $\{y_j\}_{j=1}^N$ extracted from the probe image by Eqs. (8) and (9), altogether, we can get $N$ label predictions $\{c_j\}_{j=1}^N$. Finally, we apply a simple majority-based voting strategy on $\{c_j\}_{j=1}^N$ to get the final label prediction $l$ for the probe range image.

![Fig. 4. Flowchart of MDCR+LH_STCC, the proposed method used for 3D ear or palmprint identification.](image)

Till now, we have presented our proposed universal method for classifying 3D ears or palmprints, which uses MDCR as the classification framework and L-H_STCC as the feature extraction scheme. In the following, our proposed method will be referred as MDCR+LH_STCC for short. The overall flowchart of MDCR+LH_STCC is shown in Fig. 4, using 3D ear identification as a concrete example.
5 Experiments

5.1 Datasets and the test protocol

Experiments were conducted on two benchmark 3D biometrics datasets, one for 3D palmprint recognition [31] and one for 3D ear recognition [32]. We use the recognition rate as the performance measure. In addition, the time cost consumed for one identification operation by each method was also evaluated. 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. 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 Matlab2013b.

5.2 3D palmprint identification

![Fig. 5. Samples of 3D palmprint ROIs in PolyU 3D palmprint dataset [31]. (a) and (b) are from one palm while (c) and (d) are from another palm.](image)

<table>
<thead>
<tr>
<th></th>
<th>recognition rate (%)</th>
<th>time cost for 1 identification (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCI [16]</td>
<td>91.88</td>
<td>9403.33</td>
</tr>
<tr>
<td>GCI [16]</td>
<td>91.87</td>
<td>9403.30</td>
</tr>
<tr>
<td>ST [16]</td>
<td>98.78</td>
<td>63275.86</td>
</tr>
<tr>
<td>LC [25]</td>
<td>91.73</td>
<td>70992.13</td>
</tr>
<tr>
<td>CR_L2 [28]</td>
<td>99.15</td>
<td>22.78</td>
</tr>
<tr>
<td>MDCR+LH_STCC</td>
<td>99.78</td>
<td>310.28</td>
</tr>
</tbody>
</table>

The PolyU 3D palmprint dataset [31] contains 8000 samples collected from 400 different palms, belonging to 200 volunteers. Among the volunteers, 136 were male and the other 64 were female. 20 samples from each of those palms were collected in two separate sessions, where 10 samples were collected in each session, respectively. The average time interval between the two sessions was one month. Sample 3D palmprint ROI images are shown in Fig. 5. In Fig. 5, (a) and (b) are from one palm while (c) and (d) are from another palm. In
our experiments, we took images collected at the first session as the gallery set and samples collected at the second session as the probe set. Under such an experimental setting, the gallery set has 400 classes and each class has 10 samples.

In order to demonstrate the superiority of our proposed MDCR+LH_STCC approach for 3D palmprint identification, several state-of-the-art methods in this field were evaluated. They include the mean curvature image (MCI) based method [16], the Gaussian curvature image (GCI) based method [16], the surface types (ST) based method [16], the local correlation (LC) based method [25], and the collaborative representation based method (CR_L2) [28].

The evaluation results are summarized in Table 2, from which we can have the following findings. At first, with respect to the classification accuracy, the proposed method MDCR+LH_STCC performs much better than the other state-of-the-art methods. It can get a 99.78% rank-1 recognition rate on the PolyU 3D palmprint dataset. Secondly, in terms of the running speed, CR_L2 runs the fastest and MDCR+LH_STCC can rank the second. Both CR_L2 and MDCR+LH_STCC can run greatly faster than the other competitors. The low speeds of MCI, GCI, and ST [16] should be attributed to the multiple translation-based matching strategy they adopted. That is, in order to account for the possible translation between the probe ROI $t$ and the gallery ROI $r$, multiple matches are performed by translating one set of feature maps in horizontal and vertical directions and the minimum of the resulting matching distances is considered to be the final matching distance between $t$ and $r$. For LC [25], it needs to compute a local correlation coefficient for every point, which makes it rather slow.

5.3 3D ear identification

The UND Collection J2 3D ear dataset [32] 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 among data samples. In addition, in some samples, ear regions are occluded with hair or ear rings. Each scan is a 640 × 480 range image. Several scan samples are shown in Fig. 6.

![Fig. 6. Samples of 3D side face scans in UND-J2 dataset [32].](image-url)
Table 3. 3D ear subsets used in our experiment

<table>
<thead>
<tr>
<th>subset index</th>
<th>#classes</th>
<th>gallery size</th>
<th>probe size</th>
<th>total samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>127</td>
<td>762</td>
<td>715</td>
<td>1477</td>
</tr>
<tr>
<td>2</td>
<td>85</td>
<td>680</td>
<td>461</td>
<td>1141</td>
</tr>
<tr>
<td>3</td>
<td>62</td>
<td>620</td>
<td>291</td>
<td>911</td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>468</td>
<td>168</td>
<td>636</td>
</tr>
</tbody>
</table>

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 [3], 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 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 3.

In experiments, MDCR+LH_STCC was compared with the classical ICP-based method and Zhang et al.’s method [15], which is one of the state-of-the-art approaches in the field of 3D ear identification. The evaluation results are listed in Tables 4 and 5. In Table 4, we list the rank-1 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 one each subset.

Table 4. Rank-1 recognition rates for 3D ear (%)

<table>
<thead>
<tr>
<th></th>
<th>subset 1</th>
<th>subset 2</th>
<th>subset 3</th>
<th>subset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICP</td>
<td>83.22</td>
<td>90.02</td>
<td>94.09</td>
<td>95.83</td>
</tr>
<tr>
<td>Zhang et al. [15]</td>
<td>83.78</td>
<td>90.67</td>
<td>94.50</td>
<td>96.43</td>
</tr>
<tr>
<td>MDCR+LH_STCC</td>
<td>95.24</td>
<td>96.75</td>
<td>98.28</td>
<td>99.40</td>
</tr>
</tbody>
</table>

Table 5. Time cost for one 3D ear identification operation (seconds)

<table>
<thead>
<tr>
<th></th>
<th>subset 1</th>
<th>subset 2</th>
<th>subset 3</th>
<th>subset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICP</td>
<td>5.356×10^9</td>
<td>3.763×10^9</td>
<td>1.876×10^9</td>
<td>1.287×10^9</td>
</tr>
<tr>
<td>Zhang et al. [15]</td>
<td>2.425</td>
<td>2.424</td>
<td>2.423</td>
<td>2.420</td>
</tr>
<tr>
<td>MDCR+LH_STCC</td>
<td>0.039</td>
<td>0.058</td>
<td>0.057</td>
<td>0.028</td>
</tr>
</tbody>
</table>

It can be seen that with respect to the classification accuracy, the proposed method MDCR+LH_STCC performs the best on all subsets. In addition, MD-
CR+LH_STCC runs much faster than the other methods evaluated. Particularly, the computational burden of ICP is extremely heavy, making it not suitable for large-scale identification applications.

### 5.4 Analysis of performance improvement

From the aforementioned experimental results, it can be seen that the proposed method MDCR+LH_STCC has low computational complexity and can achieve the state-of-the-art classification accuracy in two fields, 3D palmprint identification and 3D ear identification.

| Table 6. Analysis of performance improvement for 3D palmprint identification |
|---------------------------------|-----------------|
| Method                          | Rank-1 recognition rate |
| CR_L2 (CRC_RLS+LH_ST)           | 99.15            |
| CRC_RLS+LH_STCC                 | 99.30            |
| MDCR+LH_ST                      | 99.48            |
| MDCR+LH_STCC                    | **99.78**        |

| Table 7. Analysis of performance improvement for 3D ear identification |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
|                                | subset 1 | subset 2 | subset 3 | subset 4 |
| CR_L2 (CRC_RLS+LH_ST)           | 88.53    | 92.19    | 95.53    | 98.21    |
| CRC_RLS+LH_STCC                 | 93.99    | 94.47    | 96.90    | 99.40    |
| MDCR+LH_ST                      | 90.63    | 93.71    | 96.71    | 98.21    |
| MDCR+LH_STCC                    | **95.24** | **96.75** | **98.28** | **99.40** |

Actually, the proposed MDCR+LH_STCC approach is an extension of the method CR_L2 [28], which was originally proposed for 3D palmprint identification. CR_L2 uses CRC_RLS [29] as the classification framework and the local histograms of surface types (LH_ST) as the feature. As compared to CR_L2, the novelty of MDCR+LH_STCC lies largely in two directions. First, we enriched LH_ST as LH_STCC by additionally incorporating local dominant orientation information encoded by CompCode [10]. Second, instead of using a single dictionary based classification scheme as CRC_RLS does, we proposed a novel multi-dictionary collaborative representation based classification framework, MDCR. Here we explain and demonstrate the performance improvement afforded by each new aspect of MDCR+LH_STCC. Denote by CRC_RLS+LH_ST the algorithm using CRC_RLS as the classification framework and LH_STCC as the feature. Denote by MDCR+LH_ST the algorithm using MDCR for classification and LH_ST as the feature. Their performances for 3D palmprint identification and 3D ear identification are summarized in Tables 6 and 7, respectively. We also present the results achieved by CR_L2 and MDCR+LH_STCC in tables 6 and 7 for a clear comparison.
It can be seen that the results listed in Tables 6 and 7 are quite consistent and the following conclusions can be drawn. At first, CRC\_RLS+LH\_STCC performs better than CR\_L2, indicating that as a feature extraction scheme, LH\_STCC performs better than LH\_ST. It implies that LH\_ST's discriminant capability can be strengthened by incorporating local dominant orientation information. Secondly, MDCR\_LH\_ST can get better results than CR\_L2, showing that the proposed classification framework MDCR performs better than CR\_L2. It implies that the classifier based on multiple dictionaries constructed from local blocks works better than the one based on a single global dictionary. Thirdly, to some extent, the performance enhancements brought by improvements in the feature extraction strategy and the classification strategy are “independent”; when they work together, the performance can be further boosted. That explains why MDCR\_LH\_STCC performs the best in all cases.

6 Conclusions

Developing sophisticated 3D biometrics systems have attracted much attention from researchers recently. In this work, we proposed a universal framework, MDCR\_LH\_STCC, which can be used for 3D ear and 3D palmprint identification applications both. Our contributions are mainly from two aspects. At first, we proposed a novel classification scheme, namely multi-dictionary based collaborative representation, MDCR for short. Secondly, we proposed a patch-wise statistics based feature extraction scheme, LH\_STCC, which integrates the range image’s local surface type information and local dominant orientation information. Extensive experiments conducted on two benchmark datasets show that MDCR\_LH\_STCC yields much higher recognition accuracy than the compared competing methods. In addition, its computational complexity is quite low, making it quite suitable for large-scale time-critical identification applications.

References