



3D Ear Identification Based on Sparse Representation

Lin Zhang, Zhixuan Ding, Hongyu Li, Ying Shen*

School of Software Engineering, Tongji University, Shanghai, China

Abstract

Biometrics based personal authentication is an effective way for automatically recognizing, with a high confidence, a person's identity. Recently, 3D ear shape has attracted tremendous interests in research field due to its richness of feature and ease of acquisition. However, the existing ICP (Iterative Closet Point)-based 3D ear matching methods prevalent in the literature are not quite efficient to cope with the one-to-many identification case. In this paper, we aim to fill this gap by proposing a novel effective fully automatic 3D ear identification system. We at first propose an accurate and efficient template-based ear detection method. By utilizing such a method, the extracted ear regions are represented in a common canonical coordinate system determined by the ear contour template, which facilitates much the following stages of feature extraction and classification. For each extracted 3D ear, a feature vector is generated as its representation by making use of a PCA-based local feature descriptor. At the stage of classification, we resort to the sparse representation based classification approach, which actually solves an l_1 -minimization problem. To the best of our knowledge, this is the first work introducing the sparse representation framework into the field of 3D ear identification. Extensive experiments conducted on a benchmark dataset corroborate the effectiveness and efficiency of the proposed approach. The associated Matlab source code and the evaluation results have been made publicly online available at <http://sse.tongji.edu.cn/linzhang/ear/srcear/srcear.htm>.

Citation: Zhang L, Ding Z, Li H, Shen Y (2014) 3D Ear Identification Based on Sparse Representation. PLoS ONE 9(4): e95506. doi:10.1371/journal.pone.0095506

Editor: Wang Zhan, University of Maryland, College Park, United States of America

Received: February 4, 2014; **Accepted:** March 26, 2014; **Published:** April 16, 2014

Copyright: © 2014 Zhang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work is supported by the Natural Science Foundation of China under grant no. 61201394, the Shanghai Pujiang Program under grant no. 13PJ1408700, and the Innovation Program of Shanghai Municipal Education Commission under grant no. 12ZZ029. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: yingshen@tongji.edu.cn

Introduction

In modern society, with the development of hardware and the awareness of security among people, recognizing the identity of a person with high confidence has become a key issue and a topic of intense study [1]. Conventional card based, or user name and password based, authentication systems have shown unreliability and inconvenience to some extent. Instead, biometrics based methods, which use unique physical or behavioral characteristics of human beings, are drawing increasing attention in both academic research and industrial applications because of their high accuracy and robustness in the modern e-world. With the rapid development of computing techniques, in the past several decades or so, researchers have exhaustively investigated a number of different biometric identifiers, including fingerprint, face, iris, palmprint, hand geometry, voice, gait, etc [2].

Compared with classical biometric identifiers such as fingerprint [3] and face [4], the ear is relatively a new member in the biometrics family and has recently received some significant attention due to its non-intrusiveness and ease of data collection. As a biometric identifier, the ear is appealing and has some desirable properties such as universality, uniqueness and permanence [5,6]. The ear has a rich structure and a distinct shape which remains unchanged from 8 to 70 years of age as determined by Iannarelli in a study of 10,000 ears [5]. As pointed out by Chang *et al.* [7], the recognition using 2D ear images has a comparable discriminative power compared with the recognition using 2D face images.

According to the types of input data, ear recognition problems can be classified as 2D, 3D, and multimodal 2D plus 3D. Most existing studies made use of only 2D profile images and the representative works belonging to this category can be found in [7–14]. Besides the traditional 2D ear sensing, there now also exists technologies to acquire 2D plus 3D ear data simultaneously. With the developing and popularizing of 3D sensor technology, there is a rising trend to use 3D sensor instead of traditional 2D camera in ear recognition research. Compared with 2D data, 3D ear data contains more information about ear shape and is not sensitive to illumination and occlusion. In [15], Yan and Bowyer found that ear matching based on 3D data could achieve a higher accuracy than that making use of the corresponding 2D images.

In this paper, we address the problem of 3D ear identification. To this end, we propose a novel effective and efficient fully automatic 3D ear recognition approach based on the sparse representation framework [16]. The remainder of this paper will be arranged as follows. Related work and our contributions of this paper are described in Section 2. The proposed template-based ear detection approach is elaborated in Section 3. The feature extraction scheme and the classification approach are described in Section 4. Results of performance evaluations are presented in Section 5. Finally, Section 6 concludes the paper.

Related Work and Our Contributions

1. 3D Ear Recognition

3D ear biometrics is relatively a new research area and there have been a few studies conducted. Some prominent and

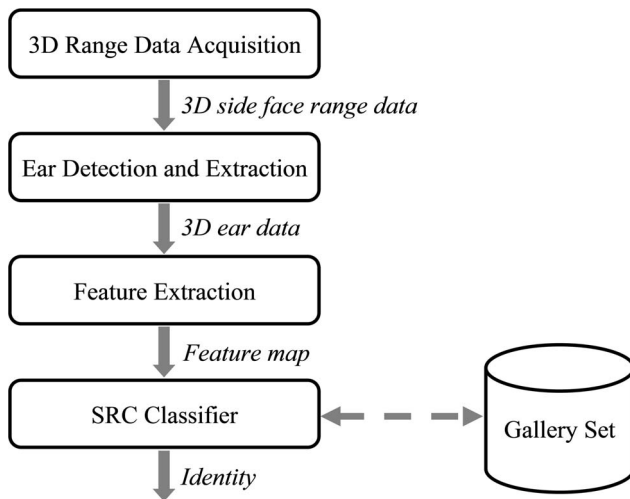


Figure 1. The general flowchart of our proposed 3D ear identification approach.

doi:10.1371/journal.pone.0095506.g001

representative relevant works are briefly reviewed here. Chen and Bhanu are among the earliest scholars conducting research on 3D ear biometrics. In [17], they proposed a method for detecting the ear region from a profile range image. Their approach is a two-step system, including model template building and online detection. The model template is obtained by averaging the shape index histograms of multiple ear samples. The online detection process consists of four steps, step edge detection and thresholding, image dilation, connected-component labeling, and template matching. In [18], Chen and Bhanu proposed a shape-model based technique for locating ears in a side face range image. However, how to match two ears was not mentioned in [18]. In [19], Chen and Bhanu proposed a two-step iterative closest point (ICP) [20] based approach for 3D ear matching. In the first step, the helix of the test ear is detected and is coarsely aligned with the model ear helix by using ICP; in the second step, the ICP algorithm iteratively refines the transformation to bring model ears and the test ear into best alignment. The root mean square (RMS) distance is used as the matching error criterion. The shortcoming of the work [19] is that all the ear regions were extracted from profile images manually. In their later work [21], Chen and Bhanu developed a fully automatic 3D ear matching framework. In [21], they proposed two shape representations for 3D ear, namely, a local surface patch (LSP) representation and a helix/antihelix representation. Both of the shape representations are used to estimate the initial rigid transformation between a gallery-probe pair. A modified ICP algorithm is then used to iteratively refine the alignment. In [22], Yan and Bowyer conducted an experimental investigation of ear biometrics and they exploited several different approaches, including the eigen-ear method using 2D intensity images as input, principal component analysis (PCA) applied to range images, Hausdorff matching of depth edge images derived from range images, and ICP-based matching of the 3D data. In their later work [23], Yan and Bowyer proposed a fully automatic 3D ear recognition system, in which for ear region detection they tried to locate the ear pit and then used the active contour algorithm [24] to extract the ear contour. With respect to the strategy for matching two ears, they again resorted to ICP. In [25], Passalis *et al.* proposed a generic annotated ear model (AEM) to register and fit each 3D ear and then a compact biometric signature was extracted containing 3D information. In their later

paper [26], they extended their work by developing a semi-automatic multi-modal 3D face and 3D ear recognition system. Such a system processes each modality separately and the final decision is determined based on the weighted average of the two similarity measures from the face and the ear modalities. In [27], Cadavid and Abdel-Mottaleb proposed an approach for 3D ear biometrics using video sequences. For each subject, a 3D ear model is derived from a video clip and ICP is adopted for computing the matching distance between two ear models. In [28], Islam *et al.* adapted the face recognition framework developed in their previous work [29] and proposed a coarse-to-fine 3D ear recognition approach. In their method, 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 range image. For ear matching, they also adopted ICP. In their latest work [30], Islam *et al.* tried to combine the 3D face and the 3D ear to build a multimodal biometrics system. In [31], Liu explored a fast recognition mechanism based on local surface matching with ICP registration to solve the 3D ear recognition problem under low cost strip point cloud environment. For a more comprehensive recent review of the ear biometrics, readers can refer to [32].

2. Sparse Representation in 3D Biometrics

It has been found that natural images can be sparsely coded by structural primitives [33] and in recent years sparse coding or sparse representation has been successfully applied to a variety of problems in computer vision and image analysis, including image denoising [34], image restoration [35,36], object classification [16,37,38], visual saliency [39], and blind image quality assessment [40]. The great success of sparse representation can be partially attributed to the progress of l_0 -norm and l_1 -norm minimization techniques [41–46].

In [16], Wright *et al.* made use of the sparse representation based classification (SRC) framework to address the face recognition problem and achieved impressive performances. With SRC, a query face image is first sparsely coded over the gallery images and then the classification is performed by checking which class yields the least coding error. Motivated by the great success of Wright *et al.*'s work, some other researchers have tried to apply SRC to other kinds of classification problems and among these works there are a few relevant to 3D biometrics. A representative work specializing on applying SRC to 3D biometrics is [47]. In [47], Li *et al.* proposed a 3D face recognition approach based on SRC. In their work, for each 3D face, a feature vector comprising a set of low-level features, such as the curvature at the vertex, the area of each triangle, the length of each edge etc., is extracted and fed into the sparse representation based classifier for classification. With a similar idea as [47], there are other two works also focusing on SRC-based 3D face recognition [48,49]. The difference among the works [47], [48] and [49] mainly lies in that they made use of different schemes for the feature extraction and selection.

3. Motivations and Contributions

From the aforementioned introduction and analysis, it can be seen that most existing 3D ear matching methods are based on ICP or its variants. While ICP is a feasible 3D matching approach for the one-to-one verification case, it is not quite appropriate for the one-to-many identification case. 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. With the number of gallery samples rising, the performance of ICP-based methods will markedly drop down.

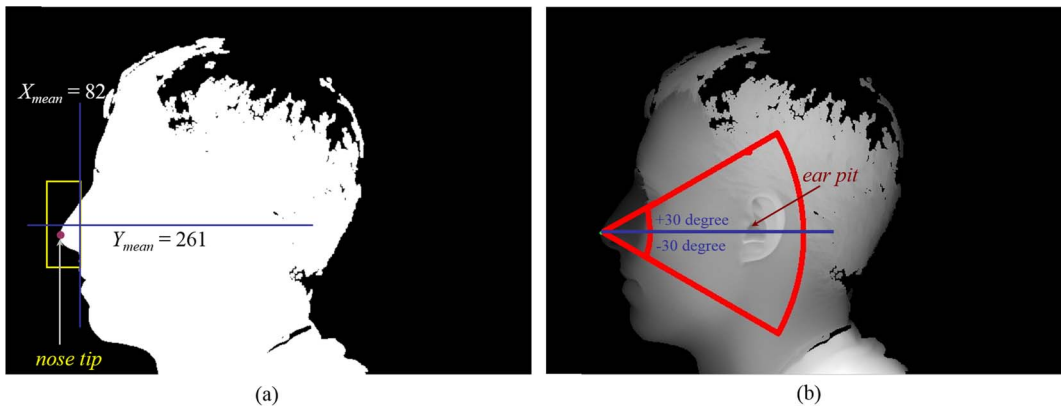


Figure 2. Illustration for the proposed ear pit detection scheme. (a) The nose tip is first located in the binary mask. (b) When the nose tip is detected, we can further locate the ear pit within a sector associated to the nose tip.
doi:10.1371/journal.pone.0095506.g002

Actually, the task of identification is essentially to find out a single class containing samples most similar to the input test sample out of the entire gallery set. To solve such a one-to-many identification

problem, Wright *et al.* [16] have demonstrated that SRC is an effective and efficient tool. In addition, researchers have also shown the feasibility of making use of SRC to solve the 3D face

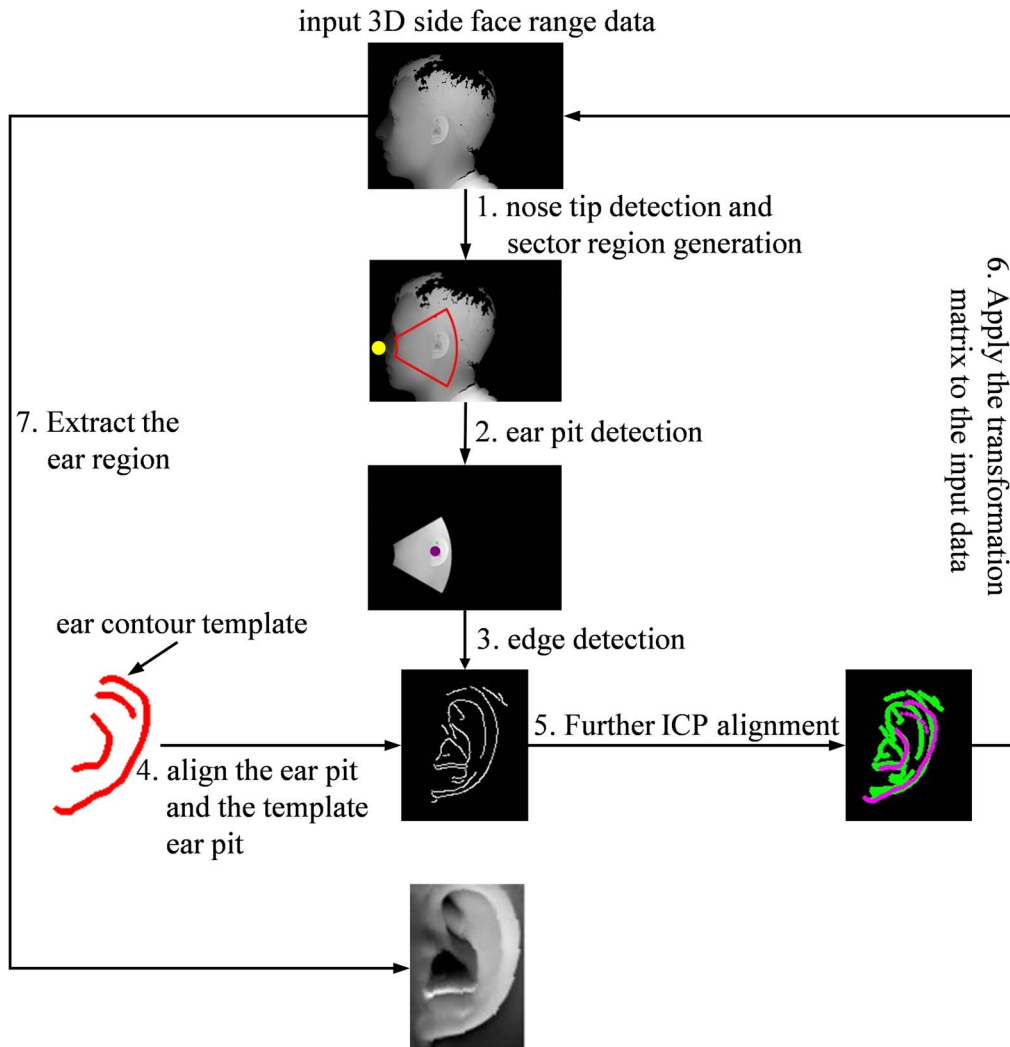


Figure 3. Flowchart of the proposed ear detection scheme.
doi:10.1371/journal.pone.0095506.g003

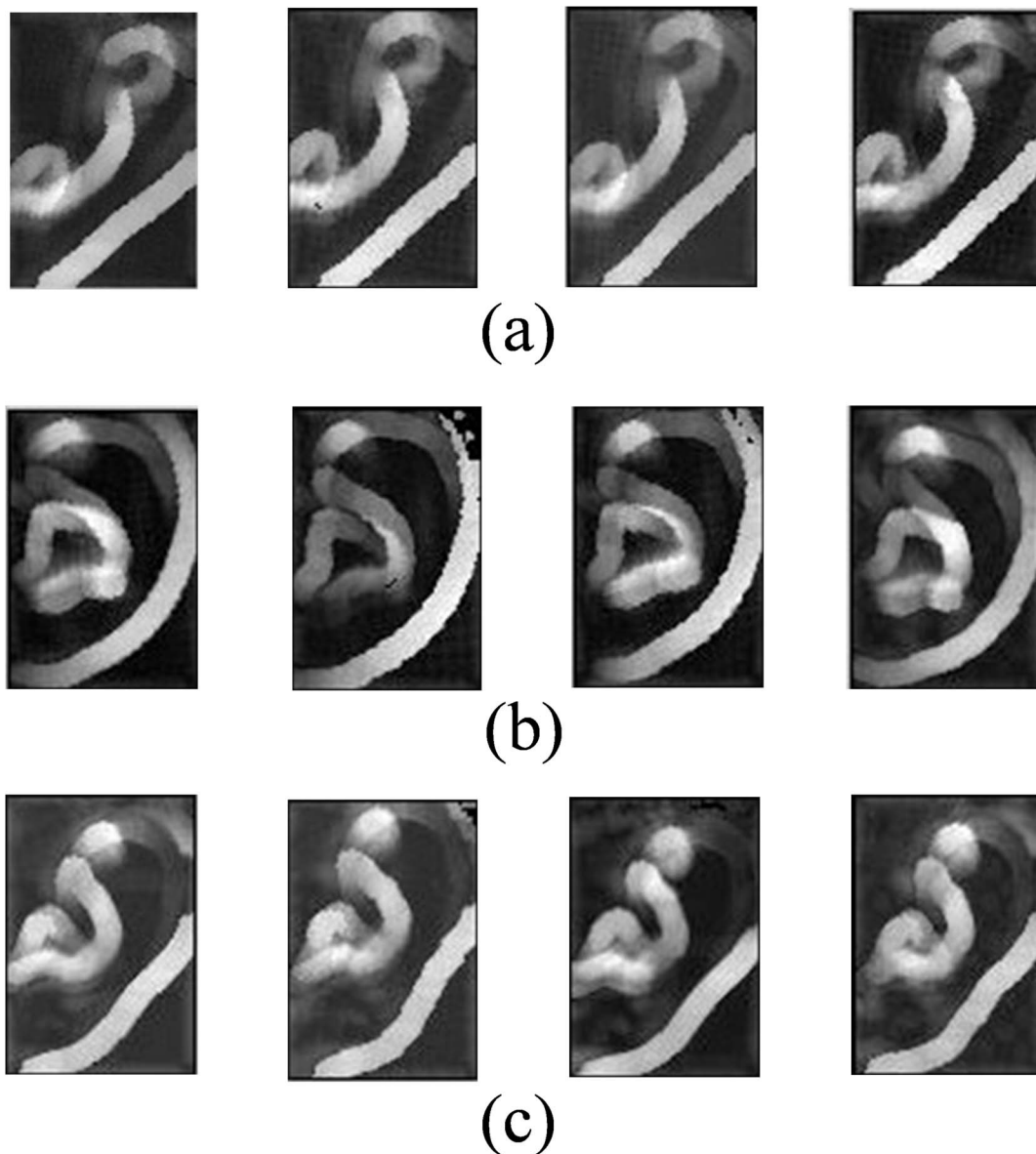


Figure 4. Feature samples computed by using the proposed feature extraction method. (a) Feature maps from different samples of ear A. (b) Feature maps from different samples of ear B. (c) Feature maps from different samples of ear C.
doi:10.1371/journal.pone.0095506.g004

recognition problem [47–49]. However, to the best of our knowledge, there is no work so far reported to apply SRC to address the 3D ear recognition problem.

Based on these considerations, in this paper, we propose a novel fully automatic 3D ear identification approach based on sparse representation, whose general architecture is shown in Fig. 1. From Fig. 1, it can be seen that our approach mainly consists of three components, ear detection and extraction, feature extraction, and SRC-based classification. To match the SRC classifier used at the classification stage, for ear detection, we propose a template based scheme which is robust to pose change. Since with such an ear detection scheme the extracted ear region will be aligned to the template ear contour model, all the extracted ears consequently reside in a common canonical coordinate system. Therefore, at the identification stage, we do not need to register the test ear sample to the gallery samples, which is crucial for the

usage of SRC. For feature extraction, we propose an effective PCA-based local descriptor for the local 3D range data, which is highly inspired by Mian *et al.*'s salient work [50]. Feature vectors are extracted from ear samples in the gallery set and they form an overcomplete dictionary \mathbf{A} . When a new test ear sample comes, its feature vector \mathbf{y} will be extracted and then its identity can be figured out by using the SRC which codes \mathbf{y} over the dictionary \mathbf{A} . Our approach takes the 3D range image as the only input and no extra color image is required. The performance of the proposed approach is evaluated on the benchmark dataset and is compared with the ICP based method. Efficacy and efficiency of our approach are corroborated by the experimental results.

Template-based Ear Detection

In this section, the proposed template-based ear detection method will be presented. Actually, most of the existing ear

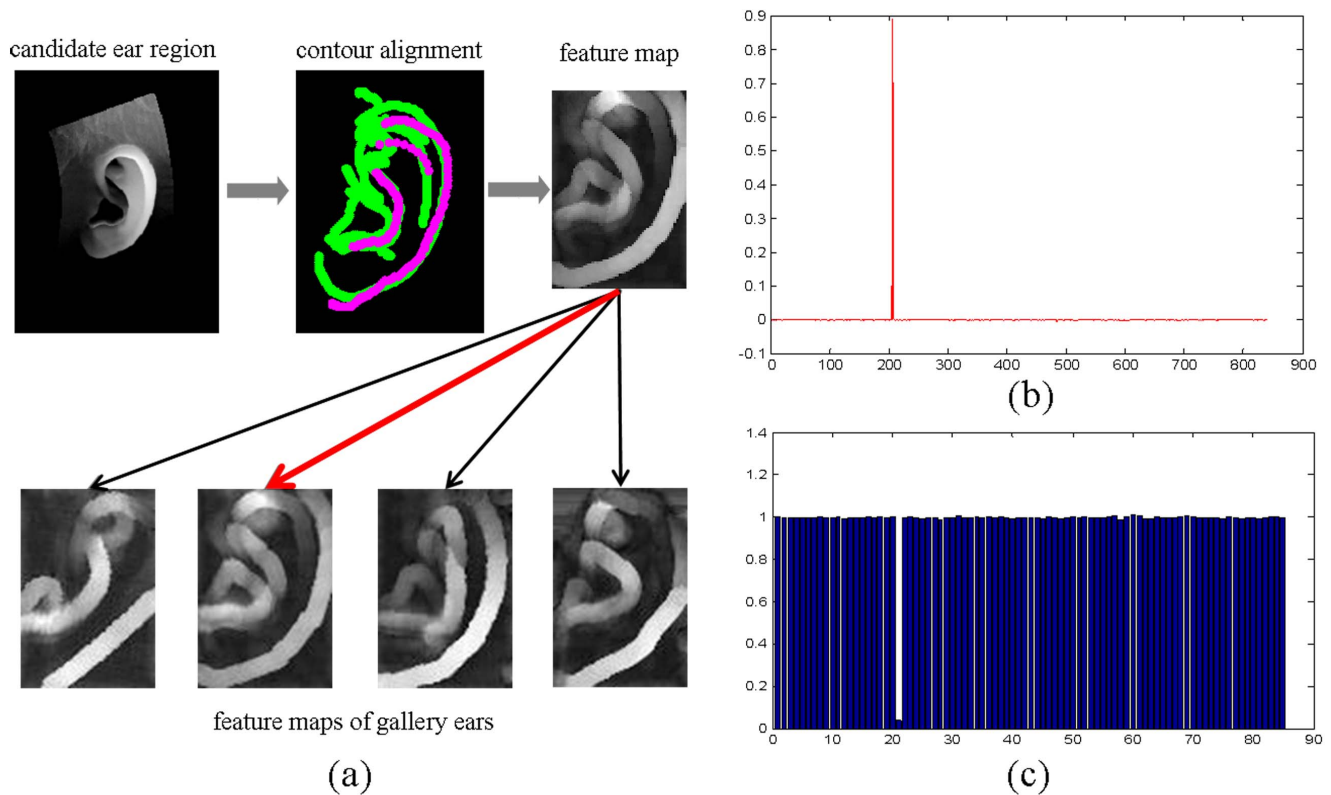


Figure 5. Illustration for the key steps of SRC-based 3D ear identification. (a) Illustration for the process of the proposed SRC based ear identification approach. (b) The values of the sparse coefficients recovered by SRC. (c) Reconstruction residuals obtained by using different ear classes to reconstruct the input test sample.
doi:10.1371/journal.pone.0095506.g005

detection or ear region extraction approaches use only 2D profile images. However, when the ear data is acquire by 3D sensors, sometimes the corresponding 2D color images cannot be gained or have occlusion problems. Thus, in our paper we present an automatic ear detection approach totally based on 3D range data. In our work, however, the role of the ear detection is not only to locate and extract the ear region out of the input 3D ear data but also to align the extracted ear region to a template ear contour model. In this way, all the extracted ears are represented in a common canonical coordinate system. In our system for each ear, the input 3D ear data is just a 640×480 side face range scan. Since there are pixels which fail to record the corresponding vertices for a range scan, a binary image (usually provided along with the 3D range data) is utilized to indicate whether a pixel contains a vertex.

1. Ear Pit Detection

The binary image actually is a mask, each pixel of which indicates whether the corresponding pixel in the range scan contains a valid vertex. In order to locate the ear pit, we first need to locate the nose tip, which will greatly narrow down the searching range for the ear pit in the following process. Identifying the location of the nose tip is accomplished in the image space of the binary mask, which comprises the following two steps,

- Record the X value along each row at which we first encounter a white pixel in the binary image and find the mean value X_{mean} of X values.
- Record the Y values of those rows whose starting “white” pixels have X values smaller than X_{mean} and greater than $X_{mean} - 100$. And we denote the mean of these Y values by Y_{mean} . Within a 60 pixel range above and below of the Y_{mean} , the

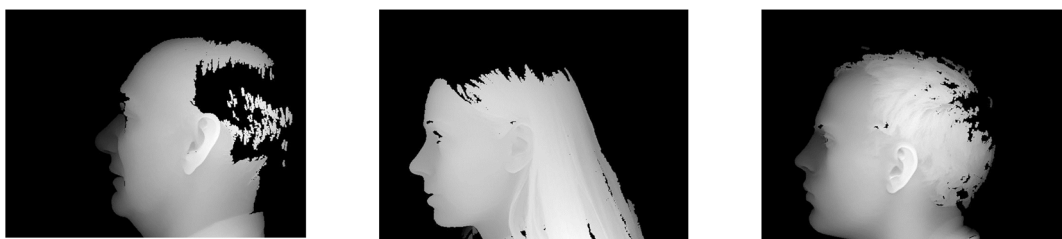


Figure 6. Samples of 3D side face range data in UNL-J2 database.
doi:10.1371/journal.pone.0095506.g006

Table 1. Rank-1 recognition rate.

	M = 5	M = 7	M = 10
ICP	83.83%	89.64%	94.09%
Our Algorithm	87.79%	91.53%	95.23%

doi:10.1371/journal.pone.0095506.t001

valid point with minimum X value is the nose tip and its position is represented by $(X_{NoseTip}, Y_{NoseTip})$.

We use the Z value in the range scan as the intensity for each pixel to generate a 2D intensity image, and the further locating of ear pit is performed in such an image. Based on the location of the nose tip, we can generate a corresponding sector which the ear pit should fall in. Such a sector can narrow down the searching range for identifying the ear pit. With the point $(X_{NoseTip}, Y_{NoseTip})$ as the center, we define this sector as a fan region spanning ± 30 degrees from the horizontal. In this sector, we reject pixels of which the corresponding vertices have a distance larger than 16 cm or smaller than 4 cm to the nose tip. To identify the ear pit, we propose a simple yet effective method. We assume that the ear pit should be the point of which the Z value is the lowest within a circular range. Following this rule, we pre-generate a random sample pattern which consists of 150 points within a circle whose radius is 30-pixels. For each point falling in the sector, we sample the points in its neighborhood via the pre-generated sample pattern. The points with the lowest depth values in their local neighborhoods are valid candidates for the ear pit. By observing the distribution of these candidates, we found that the true ear pit must lie on a cluster which contains several candidates. Hence, we remove the isolated candidates having no other candidates around, which could be caused by noise or hair interference. From the remaining candidates, we regard the one with the smallest depth value as the ear pit. Procedures for the ear pit detection are illustrated by using an example in Fig. 2.

2. Ear Contour Alignment and Ear Region Extraction

In order to partially solve the pose change problem and to make the extracted ear regions reside in a common canonical coordinate system, we adopt the ICP algorithm to align the ear contour of the sample being processed to an ear contour template created offline manually. Such an idea is inspired by the work [21]; however, there are some differences. At first, in our case, since the ear pit has already been detected, the computational cost of ICP-based contour matching could be reduced by aligning the ear pits first. Secondly, in our case, ICP matching is performed in the 2D image space, which is much faster than the one working in the 3D space.

We built an ear contour template by manually selecting the ear pit point, 30 helix points, and 10 antihelix points from one instance of samples from UND-J2 ear database [51]. The ear contour template finally generated is shown in Fig. 3. For each 3D side face image being processed, we at first extract the sector

region containing the ear. Then, we apply a Canny edge detector on the extracted sector of the depth image and we can consequently get an edge map which can be regarded as the contour of the ear. After that, we register the obtained edge map to the ear contour template roughly by aligning the detected ear pit to the template ear pit. Then, an ICP algorithm is applied in the image space to refine the alignment further. After that, we transform the original input 3D range data into the coordinate system defined by the ear contour template by using the translation and rotation matrices obtained in the ICP alignment. Finally, a pre-defined rectangular region is extracted from the transformed range data and it is taken as the resulting ear region.

By using the template-based contour alignment, the final extracted 3D ear region is robust to ear pose change and there is little difference for the ears extracted from the same person. In addition, with such a scheme, all the extracted ear regions are represented in the same canonical coordinate system defined by the ear contour template, which facilitates the following feature extraction and classification. Fig. 3 illustrates the main steps involved in our proposed ear detection scheme.

Ear Identification based on Sparse Representation

In this section, we will present our SRC-based ear identification approach in detail. At first, how to extract features from 3D ear range data will be introduced. Then, details for classification will be presented.

1. Ear Feature Extraction

In order to make use of the SRC for classification, we need to map the extracted 3D ear data into a feature vector of a fixed length. The commonly used features for 3D shapes, like Li *et al.* adopted in [47], are point curvature, triangle area, and so on. But the difficulty is to build one-to-one correspondences between two 3D shapes. In our case, however, since we have aligned all 3D ear data to a canonical coordinate system defined by the ear contour template in the ear detection step, the one-to-one correspondences have already been built automatically and the feature extraction could be directly applied to the extracted 3D ear data.

Instead of using traditional features like curvatures, we proposed a local PCA-based feature descriptor greatly inspired by Mian *et al.*'s work [50]. The extracted 3D ear could be represented as a point cloud: $\mathbf{E} = \{[x_i, y_i, z_i]^T, \text{ where } i = 1, \dots, n\}$. For each point p_i in \mathbf{E} , let $\mathbf{L} = \{[x_j, y_j, z_j]^T, \text{ where } j = 1, \dots, n_i\}$ be the points in a region cropped by a sphere of radius r centered at the point p_i .

Table 2. Time cost for 1 identification operation (seconds).

	N = 850	N = 889	N = 925
ICP	127.45	131.61	144.31
Our Algorithm	0.041	0.044	0.047

doi:10.1371/journal.pone.0095506.t002

Given the dataset \mathbf{L} , we can calculate its mean vector \mathbf{m} and its covariance matrix \mathbf{C} ,

$$\mathbf{m} = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{L}_j \quad (1)$$

$$\mathbf{C} = \frac{1}{n_i} \sum_{j=1}^{n_i} (\mathbf{L}_j \mathbf{L}_j^T - \mathbf{m} \mathbf{m}^T) \quad (2)$$

where \mathbf{L}_j represents the coordinate of the j^{th} point in the point set \mathbf{L} . Then, an eigen-value decomposition is performed on the covariance matrix \mathbf{C} and we can get a matrix \mathbf{V} comprising eigenvectors and a diagonal matrix \mathbf{D} of eigenvalues of \mathbf{C} ,

$$\mathbf{C} \mathbf{V} = \mathbf{V} \mathbf{D} \quad (3)$$

Then, we can map point set \mathbf{L} to \mathbf{L}' by,

$$\mathbf{L}'_j = \mathbf{V}^T (\mathbf{L}_j - \mathbf{m}) \quad (4)$$

Let \mathbf{L}'_x and \mathbf{L}'_y be the projections of points \mathbf{L} on its first and second principal components, respectively (these two principal components correspond to the first two largest eigenvalues). The feature value at point p_i could be defined as,

$$\delta_i = \max(\mathbf{L}'_x) - \min(\mathbf{L}'_x) - (\max(\mathbf{L}'_y) - \min(\mathbf{L}'_y)) \quad (5)$$

According to Mian *et al.* [50], δ_i actually represents the difference between the lengths of the first two principal axes of the local region around the point p_i and it will be zero if the point cloud \mathbf{L} is planar or spherical. However, if there is unsymmetrical variation in \mathbf{L} , then δ_i will have a non-zero value proportional to the variation. Such a PCA-based local feature descriptor was originally proposed for key-points detection [50]. In [50], Mian *et al.* regard the points with high δ_i values as key points. In our case, however, we simply utilize δ_i as a local feature descriptor. After computing $\{\delta_i\}$ for all the points $\{p_i\}$ in the ear region, we can obtain a feature map. In order to be used by the following classification stage, the feature map needs to be reformulated as a column vector. Fig. 4 demonstrates some examples of the calculated feature maps, shown in image format. The three rows of Fig. 4 correspond to three different ears while for each row, the four feature maps are computed from four different samples collected belonging to the same ear. From Fig. 4, we can have the following findings. At first, the proposed feature can reflect anatomical information of the ear very well. Secondly, by using the proposed method, feature maps derived from the different samples of the same ear look quite similar to each other while the ones computed from different ears are apparently different.

The dimension of the feature vectors obtained with the above scheme is quite high and it is not effective to directly feed them into the SRC for classification. Therefore, a further dimension reduction operation is necessary and to this end, we adopt the random projection approach introduced by Wright *et al.* in [16], which is quite simple yet effective.

2. Ear Recognition via SRC

Given an ear gallery set, we compute a feature vector \mathbf{v} from each ear in gallery and form them to a dictionary matrix $\mathbf{A} = [\mathbf{v}_{11}, \dots, \mathbf{v}_{1k}, \mathbf{v}_{21}, \dots, \mathbf{v}_{2k}, \dots, \mathbf{v}_{n1}, \dots, \mathbf{v}_{nk}] \in \mathbb{R}^{m \times nk}$, m here represents the feature dimension, n represents the number of ears, and k represents the number of samples for each ear in the gallery.

Given a query ear sample, denote by \mathbf{y} its feature vector. The recognition problem can be viewed as solving the following over-completed linear equation,

$$\mathbf{x}_0^* = \arg \min \|\mathbf{x}\|_0, \quad \text{subject to } \mathbf{y} = \mathbf{A} \mathbf{x} \quad (6)$$

$\mathbf{x}^* 0$ here is a l_0 -minimization solution for this equation. However, the problem of identifying the sparsest solution of an under-determined system of linear equations is NP-hard and difficult even to approximate. Fortunately, recent development in optimization theories reveals that if the solution $\mathbf{x}^* 0$ sought is sparse enough, it can be well approximated by the solution of the following l_1 -minimization problem [41],

$$\mathbf{x}_1^* = \arg \min \|\mathbf{x}\|_1, \quad \text{subject to } \mathbf{y} = \mathbf{A} \mathbf{x} \quad (7)$$

With $\mathbf{x}^* 1$, we can compute the reconstruction residual when using the samples of the ear class i to approximate the test sample \mathbf{y} as,

$$r_i = \left\| \mathbf{y} - \sum_{j=1}^k \mathbf{x}_{1ij}^* \mathbf{v}_{ij} \right\|_2 \quad (8)$$

where $\mathbf{x}^* 1_{ij}$ represents the $(ij)^{\text{th}}$ coefficient of the solution vector $\mathbf{x}^* 1$. We then can classify \mathbf{y} based on these reconstruction residuals by assigning it to the ear class that minimizes the reconstruction residual.

To better handle the noise and corruption problem, we used an extensional sparse representation by substituting $\mathbf{B} = [\mathbf{A}, \mathbf{I}]$ for the original \mathbf{A} where \mathbf{I} is the identity matrix. Therefore, the final sparse representation algorithm we adopt is,

$$\mathbf{w}_1^* = \arg \min \|\mathbf{w}\|_1, \quad \text{subject to } \mathbf{y} = \mathbf{B} \mathbf{w} \quad (9)$$

We extract $\mathbf{x}^* 1$ by decomposing $\mathbf{w}^* 1$ as $\mathbf{w}^* 1 = [\mathbf{x}^* 1, \mathbf{e}]$ and the construction residual calculation should also be substituted by,

$$r_i = \left\| \mathbf{y} - \mathbf{e} - \sum_{j=1}^k \mathbf{x}_{1ij}^* \mathbf{v}_{ij} \right\|_2 \quad (10)$$

We can easily recognize any given query ear by solving the above l_1 -minimization problem in one matching step. For solving the l_1 -minimization problem, several prominent algorithms have been developed in the past few years, including Homotopy [42], FISTA [43], DALM [44], Sparsa [45], l_1 - l_s [46], etc. In this paper, we adopt DALM algorithm [44] as the l_1 -minimization solver, which has been proved fast and accurate.

In Fig. 5(a), we use an example to demonstrate the process of the proposed SRC-based ear identification approach. After the ear region is extracted and aligned with the ear contour template, its feature map is extracted. Then, such a feature map is stacked into a feature vector and fed into the SRC for classification. From Fig. 5(b), we can see that if the identity of the test sample exists in the gallery, the solved solution is very sparse and usually, the largest coefficient corresponds to the gallery sample most similar to

the test sample. Fig. 5(c) plots the reconstruction residuals obtained by using different gallery classes to reconstruct the test sample. The least reconstruction residual usually occurs with the class having samples most similar to the test sample and thus the reconstruction residual can be an indication of the potential identity of the test sample.

Experimental Results and Discussions

In this section, we will present the evaluation results of the proposed method. The database we used in our experiment is UND collection J2 dataset [51]. The UND-J2 dataset is currently the largest 3D ear dataset which consists of 2436 side face 3D scan from 415 different persons. Several samples are shown in Fig. 6. Each 3D ear data is a 640×480 3D scan.

1. Evaluation of the Ear Detection Performance

To validate our ear detection algorithm, we manually marked ear pits for the whole UND-J2 dataset. In our experiment, we run the experiment on all the 2436 ears of UND-J2. For each 3D ear, we compared the automatic detected ear pit location ($X_{\text{auto_earpit}}$, $T_{\text{auto_earpit}}$) with the ground-truth ear pit location ($X_{\text{gt_earpit}}$, $T_{\text{gt_earpit}}$) and if the Euclidean distance of two positions were larger than 16 pixels, we regarded the ear detection for this ear as failure.

Under these experimental settings, our algorithm could achieve a 90.87% detection rate which is much higher than 85% reported in [28] by using the method proposed in [23] on the same dataset.

2. Evaluation of the Identification Performance

Although there are 415 subjects in UND-J2 database, most subjects have only 2 samples. Since the recognition based on sparse representation needs sufficient samples for each class in the gallery [16], we cannot run our experiment on the whole database.

In our experiment, we selected three subsets from UND-J2 database. The first subset contained 185 ears of UND-J2, each of which had more than 5 samples and the second subset contained 127 ears, each of which had more than 7 samples and the third subset contained 85 ears, each of which had more than 10 samples. For subset1, we randomly selected 5 samples from each ear to form the gallery and the rest of ears were formed to the test set. So the gallery size for subset1 was 925, and the test set size was 885. For subset2, we randomly selected 7 samples from each ear to form the gallery and the rest of ears were formed to the test set. Thus, the gallery size for subset2 was 889, and the test set size was 588. For subset3, we randomly selected 10 samples from each ear to form the gallery and the rest of ears were formed to the test set. Consequently, the gallery size for subset3 was 850, and the test set size was 291.

We tested our ear recognition algorithm on those three different subsets respectively and we also evaluated the performance of ICP under the same condition. Since there were multiple samples for

each ear class in the gallery, we used ICP to match a query ear with all the samples for each ear class and regarded the minimum matching error as the matching error for that ear class. Finally, we took the label of the ear class generating the minimum matching error as the identity for that query ear. Table 1 lists the rank-1 recognition rates achieved by using our algorithm and ICP, where M stands for the number of samples for each ear class in the gallery. Table 2 lists the time cost consumed by one identification operation, where N stands for the gallery size.

3. Further Discussions

From experimental results shown in Table 1, it can be seen that our proposed method performs better than ICP in terms of rank-1 recognition rate.

The greatest advantage of our algorithm over ICP is that it has a low time cost. Table 2 compares the computational time cost for one ear query process. To recognize the identity for one query ear, the ICP based algorithm has to compare the query ear to all the gallery ears and each comparison needs an ICP alignment for the two ear shapes. Without a previous rejection process, the whole recognition process will cost a lot of time. Different from ICP, the recognition process based on sparse representation just solves an l_1 -minimization problem based on pre-calculated features, so it is much faster than ICP based approaches. With gallery size rising, the computational time of ICP approach will hugely rise. However, the computational time of sparse representation based on DALM algorithm changes little with the enlargement of the gallery size, which can be reflected from the results listed in Table 2.

Conclusions

In this paper, we proposed a novel 3D ear identification approach. In order to make use of the sparse representation framework for identification, we proposed a novel template-based ear detection method. By using this method, extracted ear regions are in a common canonical coordinate system defined by the ear contour template, which highly facilitates the following feature extraction and recognition steps. Compared with the classical ICP-based ear matching methods which will match the test sample to all the gallery samples one-by-one to determine its identity, the proposed SRC-based method is more efficient. Experiments were conducted on UND-J2 3D ear database and the results indicate that the proposed method could achieve high ear detection rate, high identification accuracy, and low computational cost.

Author Contributions

Conceived and designed the experiments: LZ ZD. Performed the experiments: LZ ZD YS. Analyzed the data: LZ ZD HL. Contributed reagents/materials/analysis tools: LZ ZD HL YS. Wrote the paper: LZ YS ZD.

References

- Jain AK, Flynn PJ, Ross A (2007) Handbook of biometrics. Springer.
- Li SZ (Ed.) (2009) Encyclopedia of biometrics. Springer.
- Chen F, Huang X, Zhou J (2013) Hierarchical minutiae matching for fingerprint and palmprint identification. IEEE Transactions on Image Processing 22: 4964–4971.
- Harguess J, Aggarwal JK (2009) A case for the average-half-face in 2D and 3D for face recognition. Proceedings of the CVPR Workshops.
- Iannarelli A (1989) Ear identification. Paramount Publishing Company.
- Jain AK (1999) BIOMETRICS: Personal identification in network society. Kluwer Academic.
- Chang K, Bowyer KW, Sarkar S, Victor B (2003) Comparison and combination of ear and face images in appearance-based biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence 25: 1160–1165.
- Burge M, Burger W (2000) Ear biometrics in computer vision. Proceedings of the ICPR, 822–826.
- Hurley DJ, Nixon MS, Carter JN (2005) Force field feature extraction for ear biometrics. Computer Vision and Image Understanding 98: 491–512.
- Moreno B, Sanchez A, Velez JF (1999) On the use of outer ear images for personal identification in security applications. Proceedings of the ICCST, 469–476.
- Choras M (2005) Ear biometrics based on geometrical feature extraction. Electronic Letters on Computer Vision and Image Analysis 5: 84–95.

12. Bustard JD, Nixon MS (2010) Toward unconstrained ear recognition from two-dimensional images. *IEEE Transactions on Systems, Man, and Cybernetics-Part A* 40: 486–494.
13. Yuan L, Mu Z (2012) Ear recognition based on local information fusion. *Pattern Recognition Letters* 33: 182–190.
14. Kumar A, Chan TT (2013) Robust ear identification using sparse representation of local texture descriptors. *Pattern Recognition* 46: 73–85.
15. Yan P, Bowyer KW (2005) Ear biometrics using 2D and 3D images. *Proceedings of the CVPR Workshops*.
16. Wright J, Yang A, Ganesh A, Sastry S, Ma Y (2009) Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31: 210–227.
17. Chen H, Bhanu B (2004) Human ear recognition from side face range images. *Proceedings of the ICPR*, 574–577.
18. Chen H, Bhanu B (2005) Shape model-based 3D ear detection from side face range images. *Proceedings of the CVPR Workshops*.
19. Chen H, Bhanu B (2005) Contour matching for 3D ear recognition. *Proceedings of the IEEE Workshop Application of Computer Vision*, 123–128.
20. Besl P, McKay ND (1992) A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14: 239–256.
21. Chen H, Bhanu B (2007) Human ear recognition in 3D. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29: 718–737.
22. Yan P, Bowyer KW (2005) Empirical evaluation of advanced ear biometrics. *Proceedings of the CVPR Workshops*.
23. Yan P, Bowyer KW (2007) Biometric recognition using three-dimensional ear shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29: 1297–1308.
24. Kass M, Witkin A, Terzopoulos D (1988) Snakes: Active contour models. *International Journal of Computer Vision* 1: 321–331.
25. Passalis G, Kakadiaris IA, Theoharis T, Toderici G, Papaioannou T (2007) Towards fast 3D ear recognition for real-life biometric applications. *Proceedings of the AVSS*, 39–44.
26. Theoharis T, Passalis G, Toderici G, Kakadiaris IA (2008) Unified 3D face and ear recognition using wavelets on geometry images. *Pattern Recognition* 41: 796–804.
27. Cadavid S, Abdel-Mottaleb M (2008) 3D ear modeling and recognition from video sequences using shape from shading. *IEEE Transactions on Information Forensics and Security* 3: 709–718.
28. Islam SM, Davies R, Bennamoun M, Mian AS (2011) Efficient detection and recognition of 3D ears. *International Journal of Computer Vision* 95: 52–73.
29. Mian A, Bennamoun M, Owens R (2008) Keypoint detection and local feature matching for textured 3D face recognition. *International Journal of Computer Vision* 79: 1–12.
30. Islam SM, Davies R, Bennamoun M, Owens RA, Mian AS (2013) Multi-biometric human recognition using 3D ear and face features. *Pattern Recognition* 46: 613–627.
31. Liu H (2013) Fast 3D ear recognition based on local surface matching and ICP registration. *Proceedings of the INCoS*, 731–735.
32. Abaza A, Ross A, Hebert C, Harrison MAF, Nixon MS (2013) A survey on ear biometrics. *ACM Computing Surveys* 45: (39: 1–33).
33. Olshausen BA, Field DJ, Sparse coding with overcomplete basis set: A strategy employed by V1? *Vision Research* 37: 3311–3325.
34. Elad E, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing* 54: 3736–3745.
35. Mairal J, Bach F, Ponce J, Sapiro G (2010) Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research* 11: 19–60.
36. Yang J, Wright J, Huang T, Ma Y (2008) Image superresolution as sparse representation of raw patches. *Proceedings of the CVPR*, 1–8.
37. Yang J, Yu K, Gong Y, Huang T (2009) Linear spatial pyramid matching using sparse coding for image classification. *Proceedings of the CVPR*, 1794–1801.
38. Wagner A, Wright J, Ganesh A, Zhou Z, Mobahi H, Ma Y (2012) Toward a practical face recognition system: robust alignment and illumination by sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34: 372–386.
39. Yang J, Yang M (2012) Top-down visual saliency via joint CRF and dictionary learning. *Proceedings of the CVPR*, 2296–2303.
40. He L, Tao D, Li X, Gao X (2012) Sparse representation for blind image quality assessment. *Proceedings of the CVPR*, 1146–1153.
41. Donoho D (2006) For most large underdetermined systems of linear equations the minimal l_1 -norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics* 59: 797–829.
42. Malioutov D, Cetin M, Willsky A (2005) Homotopy continuation for sparse signal representation. *Proceedings of the ICASSP*, 733–736.
43. Beck A, Teboulle M (2009) A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* 2: 183–202.
44. Yang J, Zhang Y (2011) Alternating direction algorithms for l_1 -problems in compressive sensing. *SIAM Journal on Scientific Computing* 33: 250–278.
45. Wright SJ, Nowak RD, Figueiredo MAT (2009) Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing* 57: 2479–2493.
46. Kim SJ, Koh K, Lustig M, Boyd S, Gorinevsky D (2007) An interior-point method for large-scale l_1 -regularized least squares. *IEEE Journal on Selected Topics in Signal Processing* 1: 606–617.
47. Li X, Jia T, Zhang H (2009) Expression-insensitive 3D face recognition using sparse representation. *Proceedings of the CVPR*, 2575–2582.
48. Tang H, Sun Y, Yin B, Ge Y (2011) 3D face recognition based on sparse representation. *Journal of Supercomputing* 58: 84–95.
49. Huang D, Ouji K, Ardabilian M, Wang Y, Chen L (2011) 3D face recognition based on local shape patterns and sparse representation classifier. *Proceedings of the MMM*, 206–216.
50. Mian AS, Bennamoun M, Owens R (2008) Keypoint detection and local feature matching for textured 3D face recognition. *International Journal of Computer Vision* 79: 1–12.
51. CVRL Datasets. Available: <http://www3.nd.edu/~cvrl>. Accessed 2014 Mar 30.