

Article

Local and Holistic Feature Fusion for Occlusion-Robust 3D Ear Recognition

Qinping Zhu  and Zhichun Mu *

School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 10083, China; zqp_2000@126.com

* Correspondence: mu@ies.ustb.edu.cn; Tel.: +86-10-62334995

Received: 1 September 2018; Accepted: 29 October 2018; Published: 1 November 2018



Abstract: Occlusion over ear surfaces results in performance degradation of ear registration and recognition systems. In this paper, we propose an occlusion-resistant three-dimensional (3D) ear recognition system consisting of four primary components: (1) an ear detection component, (2) a local feature extraction and matching component, (3) a holistic matching component, and (4) a decision-level fusion algorithm. The ear detection component is implemented based on faster region-based convolutional neural networks. In the local feature extraction and matching component, a symmetric space-centered 3D shape descriptor based on the surface patch histogram of indexed shapes (SPHIS) is used to generate a set of keypoints and a feature vector for each keypoint. Then, a two-step noncooperative game theory (NGT)-based method is proposed. The proposed symmetric game-based method is effectively applied to determine a set of keypoints that satisfy the rigid constraints from initial keypoint correspondences. In the holistic matching component, a proposed variant of breed surface voxelization is used to calculate the holistic registration error. Finally, the decision-level fusion algorithm is applied to generate the final match scores. Evaluation results from experiments conducted show that the proposed method produces competitive results for partial occlusion on a dataset consisting of natural and random occlusion.

Keywords: 3D ear recognition; occlusion; local feature matching; holistic feature matching; noncooperative game theory

1. Introduction

The modalities of the ear offer distinct advantages when applied to human recognition, as they involve a wealth of permanent structural features that retain the same shape and structure from the age of eight to seventy and remain invariable under various expressions [1]. Contemporary research on ear recognition has explored the possibility of using both two-dimensional (2D) and three-dimensional (3D) images and models of the ear for recognition. Although 2D ear recognition is widely used owing to its convenience and the efficiency of 2D image acquisition [2], the surface morphologies of many 2D images from a subject differ significantly for different poses and illumination conditions, and only the information projected onto the plane of the camera lens can be obtained. In contrast, 3D ear data are more robust to pose and illumination conditions and can provide detailed and rich depth information about the anatomical structure of the ear [3]. Automatic recognition methods do not require human intervention throughout the recognition process and are thus convenient for exploitation. In the ear-recognition domain, some recent studies have focused on automatic ear recognition using 3D range images, as they are robust against imaging conditions, contain surface shape information related to the anatomical structure of the ear, and are impervious to environmental illumination [4]. Various studies have also been conducted on 3D ear recognition with noise and virtually no occlusion. In the context of feature representation, automatic 3D ear recognition can be categorized as follows:

holistic feature-based recognition (e.g., an improved iterative closest point (ICP) method for 3D shape matching [5]); local feature-based recognition (e.g., an integrated local surface patch algorithm [6]); and multi-feature fusion-based recognition such as approaches based on local and holistic feature fusion [3] and geometrical feature- and depth-information-based indexing [7]. However, when occlusion is present, the majority of 3D ear registration and recognition algorithms fail to provide accurate ear point correspondences owing to the surface points being occluded. The resulting alignment between ear surfaces is usually incorrect, leading to low recognition rates.

Only a few researchers have investigated the occluded automatic 3D ear recognition problem. To the best of our knowledge, the best recognition rate is presented in [8], where the similarity of the components of each ear pair is computed on the basis of a modified ICP using the local surface variation (ICP-LSV) algorithm and the location of the ear pit. Related experimental results have shown that this method can achieve recognition accuracies of 94.2% and 97.6% when the ear is respectively obscured by minor amounts of hair and jewelry in relatively fixed positions [8]. However, when the ear image contains randomly positioned larger occlusions, the ear pit may not be visible, and the method becomes susceptible to paralysis.

Several types of NGT-based object matching methods [9,10] have been proposed owing to the robust advantages for rigid 3D objects in a complex and chaotic scene. With the NGT framework, these methods can choose a strategy set corresponding to the maximum earnings from a set of candidate point pairs that contain a large number of outliers. The chosen set can contain as many keypoint correspondences as possible to satisfy the rigid matching requirements. In experiments conducted, we observed that the object recognition scheme reported in the literature [9] exhibited suboptimal performance. The reasons for this phenomenon are as follows. When suffering from partial occlusion, the selection of the initial strategy set is too dependent on associating a small number of neighbor scene local descriptors with the model local descriptor, which can result in some true keypoint correspondences being overlooked. Furthermore, isolated use of the number of true keypoint correspondences generated by a local feature as a score for ear recognition yields suboptimal recognition rates. Our experiments illustrate these phenomena. In this paper, to effectively recognize ears with partial occlusion, we propose an automatic 3D ear recognition system that combines an SPHIS local descriptor, a two-step NGT-based local matching method, and a breed surface voxelization-based holistic matching method. The system diagram of our local and holistic matching fusion method is shown in Figure 1. Our motivation is as follows. The SPHIS local descriptor captures information from each local surface. In contrast, the two-step NGT-based matching algorithm can filter the outlier correspondence by enforcing a rigid constraint on the keypoint correspondence and estimating the occlusion rate of the local surface patch enclosing every keypoint. When combined effectively, they can provide complementary information describing the 3D ear structure and jointly enhance the local identification performance. Furthermore, the local matching method has been found to be more robust to occlusion and clustering, whereas the holistic matching method can obtain the registration error based on the entire surface without excluding any information that describes the ear, except the outlier correspondence. They can provide complementary knowledge to jointly enhance the matching performance, even if the relevant subject contains partial occlusion.

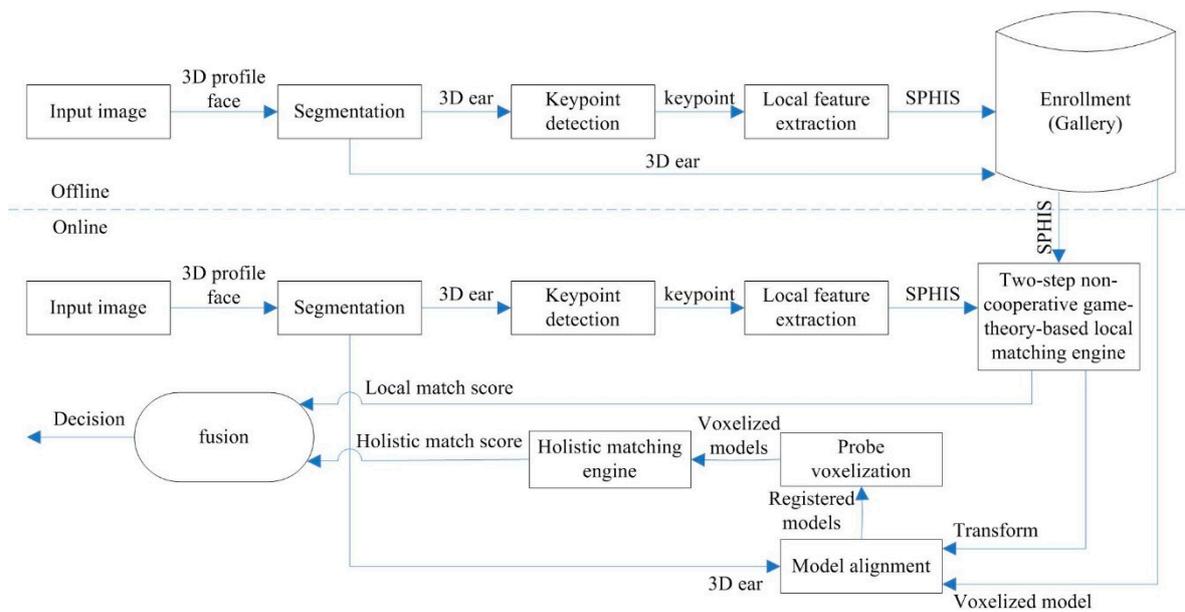


Figure 1. System diagram of the proposed 3D ear recognition system.

Identity verification based on biometrics is increasingly being used in everyday life. Ear verification can be regarded as a one-to-one ear recognition task. As the proposed method can robustly match a non-occluded gallery ear image with a partially occluded probe ear image, it can effectively solve the verification problem of partially occluded ear images. Because conducting a one-to-one ear recognition experiment after an ear recognition experiment is redundant, this paper does not present the ear verification in the experimental stage. Interested readers can evaluate the verification performance of the proposed method with reference to the evaluation method described in [11].

Below, Section 2 reviews related work and addresses the contributions of our method. Section 3 describes the automatic segmentation of the 3D ear for a given profile range image. Section 4 outlines the proposed techniques for local feature matching, including the keypoint detection method and similarity-transformation-invariant local surface patch representation, and describes the two-step NGT-based matching method. Section 5 describes the variant of the breed voxelization method for holistic matching. Section 6 presents the fusion structure based on the match scores generated by the local and holistic matching components. Section 7 provides experimental results that highlight the performance of the proposed system in terms of natural and random occlusion. Finally, the conclusions of this study and the plans for future research are presented in Section 8.

2. Related Work and Contribution

The proposed approach uses a 3D range image for automatic ear recognition and coregistered 2D ear images for automatic detection. To robustly recognize the ear, the proposed technique uses a proposed two-step NGT-based matching method that generates a correspondence that maintains a rigid constraint. In the following subsections, we briefly outline the relevant research in 2D ear detection, automatic ear recognition methods that only use a 3D range image, and recognition approaches based on NGT along with their limitations, and discuss the contributions of our method.

2.1. Automatic 2D Ear Detection

In this section, we summarize the different aspects of relevant published studies. Most automatic 2D ear detection approaches depend on the mutual morphological properties of the ear, for instance, the characteristic edges [12], model matching [13], and learning algorithms [14].

Numerous characteristic-edge-based ear detection methods have been proposed, e.g., the force-field transform approach for ear detection [15], canny edge-detector-based ear detection [12], a tracking

method that combines both a skin-color model and intensity contour information for ear detection [16], Hough transform (HT)-based ear detection [17], ear detection based on an image ray transform method [18], ear detection based on the connected components in a graph [19], pit detection and an active contour algorithm for extracting the ear [20], ear detection based on a breed active contour model [21], and the entropic binary particle swarm optimization approach for ear localization [22].

Several modeling matching-based detection methods have been published, e.g., shape-based localization [23], and a modified Hausdorff distance-based ear localization method [13].

Learning-based detection methods have been proposed owing to their accuracy and robustness advantages for ear detection, e.g., the AdaBoost algorithm and its modified version [24,25] and ear detection involving faster region-based convolutional neural network (Faster R-CNN) frameworks [26], a modified multiple scale faster region-based convolutional neural network [14], geometrics morphometrics and deep learning [27] and convolutional encoder-decoder networks for ear detection [28]. Automatic ear detection methods have achieved good performance under no occlusion or noise conditions; for example, the accuracy of the method in [12] is 93.34% for 700 images, and that in [19] is 97.57% for 267 images. The accuracies of the methods in [17] and [18] are 100% and 98.4%, respectively, for 252 images from the XM2VTS database. The detection accuracies of the method in [20] are 78.8% and 85.5% for 415 images, that in [24] is 100% for 203 images, and that in [14] is 100% for the UND-J2 database from the UND database. The method in [21] achieved an accuracy rate of 76.43% for 700 images from the ColorFERET database, and the best accuracy rate of the method in [22] for 240 images from the CMU PIE database was 92.92%. The method in [23] was tested using a self-established dataset consisting of 212 profile images and realized an accuracy rate of 96.2%. The method in [25] can achieve a slightly higher detection hit rate and much lower detection false-alarm rate than the former AdaBoost algorithm [24] for a self-built portrait database. Although the ear is partially occluded, the performance of ear detection approaches decreases; for example, the method in [17] was tested with 252 images, each with occlusion rates of 40% and 50%, and achieved accuracies of 83% and 66%, respectively. The detector failed in the case where the majority of the ear is occluded [24], and the method in [14] achieved an accuracy rate of 94.7% for a self-built dataset containing many images, each with an occlusion rate greater than 60%. In [28], average IoU values around 50% were achieved for difficult images with variations across pose, race, and occlusion. Under conditions including occlusion, complexity, and a chaotic scene, the Faster R-CNN algorithm obtained the most competitive detection results.

2.2. Automatic 3D Ear Recognition

In this section, we follow the feature-representation-based classification criterion described in Section 1 to summarize the state-of-the-art (SOA) automatic ear recognition methods that only use 3D range images.

Yan and Bowyer [20] proposed an ear biometric system in which the ear region is detected by locating the landmarks containing the ear pit and the tip of the nose. Using the cropped region, a breed ICP algorithm is used for 3D ear recognition. Chen and Bhanu [29] proposed a completely automatic 3D ear identification strategy, in which they used two 3D shape representations of the ear—a local patch representation for the salient-featured points and a helix/antihelix representation from the ear helix/antihelix localization step. A modified ICP matching algorithm is then used. Cadavid and Abdel-Mottaleb [30] obtained 3D ear biometrics using uncalibrated video sequences. By using the shape from shading (SFS) technique, a 3D model is reconstructed on the basis of a series of video images that are registered by a variant ICP algorithm. Islam et al. [31] proposed an ear recognition system, in which the ear region is detected by using a breed AdaBoost detector. Using the detected ear region, a local feature is used to extract a region with feature-rich data points, and an ICP approach is used to match the ear. Zhang et al. [32] proposed a fully automatic 3D ear recognition system. They first detect the ear region by using the nose tip and ear pit, and subsequently generate a feature vector as each ear representation. In the final classification stage, they resort to a sparse-representation-based

classification approach. Zhang et al. [8] proposed an automatic 3D ear recognition method. First, a Faster R-CNN framework [26] is used to extract a rectangular box containing an ear image. Then, a local surface variation is used to detect the ear accurately. Finally, a one-step ICP method is applied to match the ear pair.

Chen and Bhanu [6] proposed a local-feature-based recognition method. First, the correspondence of the integrated local surface patches is extracted. Subsequently, outliers are filtered by using geometric constraints, and the discriminating power for human identification is finally evaluated. In [11] and [33], Zhou et al. proposed generalized 3D ear recognition using local and holistic feature matching. A rectangular box containing the ear region is first detected by introducing a histogram of the indexed shape (HIS) feature, and the match score is separately obtained by using the extracted local and holistic features. The final recognition result is generated on the basis of the decision-level fusion rule. Maity and Abdel-Mottaleb [7] proposed an indexing-based ear identification method, in which they use a flexible mixture model and an active contour algorithm to automatically segment the ear. Four 2D shape representations of the ear are used—specifically, round and rectangular, for categorization. Then, an SPHIS feature is used to extract a local feature, and a K-dimensional search tree (KD-tree) and a pyramid technique are finally used for ear identification.

Liu et al. [34] designed a 3D ear recognition system that combine local and holistic features. First, the 3D ear dataset is acquired by a special designed laser scanner. Subsequently, a global feature class and a local feature class are defined from a 3D ear image. Eventually, these features are optimally combined for 3D ear identification. Ganapathi et al. [35] proposed a geometric-statistics-based identification technique. To compute the descriptor, they first extracted feature keypoints by making use of surface variations and then defined a descriptor vector for every keypoint.

Under no occlusion and noise conditions, 3D ear recognition approaches have achieved satisfactory recognition performance. For example, the method in [6] achieved a rank-1 recognition rate of 90.4% for a self-built database, the method in [34] produced matching results with an equal error rate (EER) of 2.2% on a 3D ear database, the methods in [7,11,20,29,33,35] achieved a rank-1 accuracy rate greater than 96% for the corresponding UND sub-database, and the method in [30] obtained a rank-1 identification rate of 95% for many images from the West Virginia University (WVU) database. Among the existing fully automatic 3D ear recognition methods, we found that 3D ear recognition methods that combine local and holistic features [33] outperform other ear recognition methods. A possible reason for this is that these methods effectively combine two complementary features. Therefore, solving the problem of 3D ear recognition by using complementary features has become a promising research direction.

2.3. Deep-Learning-Based Recognition

In this section, we summarize the SOA ear recognition systems and 3D recognition systems based on the deep learning architecture.

Tian and Mu [36] proposed using a deep convolutional neural network (CNN) and provided a visualization of the learned network for ear identification. They presented a CNN with three convolutions and two fully connected layers and evaluated their approach on the USTB ear database. Chowdhury et al. [37] proposed an ear biometric recognition scheme that uses the local features of the ear and employs a neural network. The scheme firstly estimates the ear region from the input image and then extracts the edge features from the detected ear. Then, a neural network is used to identify the subject by matching the extracted features of the subject with a feature database.

Omara et al. [38] proposed a method that extracts deep features of ear images based on VGG-M Net to solve the ear identification problem. For computational efficiency, principal component analysis is applied to reduce the dimension. Then, pairwise SVM is used for classification. They evaluated their approach on two public ear databases: USTB I and USTB II.

Emeršič et al. [39] built a CNN-based ear recognition model. They explored different strategies towards model training with limited amounts of training data and showed that by selecting an

appropriate model architecture, using aggressive data augmentation and selective learning on existing (pre-trained) models, they were able to learn an effective CNN-based model using a little over 1300 training images. With their model, they were able to improve on the rankone recognition rate of the previous state-of-the-art by more than 25% on a challenging dataset of ear images captured from the web.

Almisreb et al. [40] applied transfer learning to the well-known AlexNet CNN for subject recognition based on ear images. They adopted and fine-tuned AlexNet CNN to suit their problem domain. To train the fine-tuned network, they allocated 250 ear images for training, and 50 ear images were used for validation and testing.

Zhang et al. [41] explored the robust CNN architecture for ear feature representation and verification. First, they replaced the last pooling layers with spatial pyramid pooling layers to fit arbitrary data sizes and obtain multi-level features. Eventually, three CNNs with different scales of ear data were assembled for user verification.

Hansley et al. [42] presented an untrained ear recognition framework. To this end, they developed CNN-based solutions for ear normalization and description. They fused learned and state-of-the-art handcrafted features to improve recognition. They presented a two-stage landmark detector that operated under untrained scenarios, and used the results generated to perform geometric image normalization that boosted the performance of all evaluated descriptors.

Ying et al. [43] proposed a deep CNN-based human ear recognition algorithm. First, a deep network structure based on CNN was designed for the ear recognition problem. Then, an optimal activation function was selected and dropout technology was introduced in the final fully connected layer to prevent network over-fitting. The network model was subsequently trained using numerous human ear image samples to determine the number of feature graphs, the setting of the learning rate, and other parameters in the network. Finally, human ear recognition tests were conducted on the trained network model.

Yaman et al. [44] analyzed in detail the extraction of soft biometric traits, age, and gender from ear images. In their study, they used both geometric features and appearance-based features for ear representation. The well-known CNN models AlexNet, VGG-16, GoogLeNet, and SqueezeNet were adopted in the study.

Fan et al. [45] proposed multi-modal recognition of human face and ear based on deep learning. First, Faster Region CNN is used to detect human faces and ears. Then, CNNs are used to train their own nets and recognize people using their faces and ears, respectively, in the images. Finally, a Bayesian decision fusion method is used to fuse the recognition results of the human face and ear to improve the recognition rate. Garcia-Garcia et al. [46] proposed VoxNet, as an improvement over the well-known approaches by integrating a volumetric occupancy grids representations with a supervised CNN framework.

Qi et al. [47] improved both volumetric CNNs and multi-view CNNs based on extensive analysis of existing approaches. In addition, they examined multi-view CNNs, where they introduced multiresolution filtering in 3D.

Xu et al. [48] introduced an efficient 3D volumetric representation method for training and testing CNNs. They also built several datasets based on the volumetric representation of 3D digits, with different rotations along the x , y , and z axes taken into account. Finally, they introduced a model based on the combination of CNN models, with structure based on the classical LeNet.

Klokov and Lempitsky [49] presented a novel deep learning structure for 3D object recognition tasks in unstructured point clouds. Their structure performs multiplicative transformations and shares parameters of these transformations based on the subdivisions of the point clouds imposed on them via Kd-trees. They do not rely on grids, and therefore avoid poor scaling behavior.

Feng et al. [50] proposed a group-view CNN (GVCNN) framework for hierarchical correction modeling towards discriminative 3D shape description. They first use an expanded CNN to extract a view level descriptor. Then, a grouping module is introduced to estimate the content discrimination of

each view, based on which all views can be split into different groups according to their discriminative level. A group level description can then be further generated by pooling from view descriptors. Finally, all group level descriptors are combined into a shape level descriptor according to their discriminative weights.

The capabilities of biometric systems have recently made extraordinary leaps as a result of the emergence of deep learning. However, in real-world applications, such as passport identification and law enforcement, where only one sample per person is usually registered in the gallery, the per person sample size required for training most of the existing ear recognition methods based on deep learning is difficult to meet [51].

2.4. Noncooperative Game Theory (NGT)-Based Recognition

NGT has been successfully applied for object recognition [9,10]. Suppose we have K gallery samples and a probe sample; the initial corresponding keypoint set between the j^{th} gallery and the probe sample is denoted as $S_j = \{g_{i,j}p_{i_k} | i = 1, 2, \dots, m_j, k = i_1, i_2, \dots, i_l\}$, where $g_{i,j}$ is the i^{th} keypoint stretch of the j^{th} gallery sample, and p_{i_k} is the k^{th} keypoint associated with $g_{i,j}$ and obtained from the probe sample. S_j is obtained via heuristic methods. Each element in S_j is treated as a strategy; thus, S_j is also considered to be a strategy set.

Let $g_{i_1,j}p_{i_{1_k}}$ and $g_{i_2,j}p_{i_{2_k}}$ respectively denote the c^{th} and d^{th} strategies from S_j , which constitute a strategy pair. For a matching rigid body, the earnings of the strategy pair e_{cd} are expressed as

$$e_{cd} = \begin{cases} 0 & p_{i_{1_k}} = p_{i_{2_k}} \text{ or } g_{i_1,j} = g_{i_2,j} \\ \frac{\min(d(g_{i_1,j}, g_{i_2,j}), d(p_{i_{1_k}}, p_{i_{2_k}}))}{\max(d(g_{i_1,j}, g_{i_2,j}), d(p_{i_{1_k}}, p_{i_{2_k}}))} & \text{otherwise} \end{cases} \quad (1)$$

where $d(i, j)$ is the distance between points i and j . For all the strategies stretched by S_j , we calculate the earnings per strategy pair in accordance with the method described above; then, all earnings construct an earnings matrix E .

We are going to play a game that consists of two players. For maximum benefit, each player should select some strategies from a strategy set. Initially, we set each strategy to occupy an almost-equal share. Then, we update the share of each strategy by applying the following replicator dynamics equation:

$$x_c(t+1) = x_c(t) \frac{(Ex(t))_c}{x(t)^T Ex(t)} \quad (2)$$

where $x(t)$ is a column vector consisting of the share of each strategy after the t^{th} iteration, and the term with the subscript c is the c^{th} element of the corresponding upper-left vector. When the occupied share of each strategy no longer changes, a Nash equilibrium is reached, and iteration is terminated. A strategy that has a greater share than a threshold value is taken as the retained correspondence. The match score is represented by N_j , which is the number of retained correspondences. $\max\{N_j | j = 1, 2, \dots, K\}$ can encode the identity of the probe sample.

2.5. Contributions

In summary, automatic 3D ear recognition methods based on local and holistic feature fusion achieve very competitive performance for 3D ears when there is virtually no occlusion. However, in the framework of local- and holistic-feature fusion-based recognition methods, every probe keypoint is strictly limited. Thus, when the probe ear is partially occluded, part of the local surface changes, inevitably resulting in some probe keypoints that cannot be matched with the ground-truth gallery keypoint or with any one gallery keypoint. Thus, the local match score is easily affected. Although NGT-based methods are effective for obtaining as many true local keypoint correspondences as possible and alignment and recognition of a 3D subject when occlusion is present, these methods that

can only match these keypoints with a local surface patch are less affected by occlusion and obtain suboptimal performance.

The main contributions of this paper are as follows. A 3D ear identification system that can automatically and robustly recognize the ear in the presence of partial occlusion, and which requires only one sample per person in the gallery, is proposed. By proposing a two-step NGT-based matching method, we first realize the fully automatic registration of a 3D ear pair for a randomly located relatively larger occlusion. A holistic matching approach based on a variant of breed surface voxelization is developed to obtain the cosine similarity through the correspondence among points and the structure of the surface of the ear under partial occlusion. Finally, we evaluate the proposed method in terms of rank-1 recognition rate and time consumption in cases with natural occlusion and random occlusion.

3. Ear Data Detection

Automatic detection without user intervention is among the most challenging problems in computer vision. The primary problem for ear detection is occlusion, which can occur because of long hair, over-the-ear headphones, earrings, or other objects.

The ear is detected in 2D profile images using the Fast-R-CNN-based detector developed by Zhang and Mu [14]. This detector was chosen because it is fully automatic, fast, and robust against occlusion and has been reported to yield an accuracy of 100% for the UND-J2 profile face database with 1800 images of 415 subjects [14]. The corresponding 3D data were then extracted from the coregistered 3D profile data, as described in [8].

The occluded ear dataset from the UND-J2 profile face database only contains two types of occlusion; hair occlusion is only near the edge of the ear, and earring occlusion only exists on the lower ear. In real applications, the position of the occlusion and the percent corrupted were neither fixed nor predictable. Thus, we simulated various levels of contiguous occlusion by substituting an unrelated image with size determined by the percentage of occlusion for a randomly located square block in a bounding box enclosing a 2D ear image. Then, the corresponding location between the camera lens and coregistered probe 3D image is occupied by the corresponding occlusion. The distance between every point contained in the occlusion and the coregistered probe 3D image is a random positive number less than the distance between the camera lens and the coregistered probe 3D image. Figure 2 shows some 2D images (first row) and coregistered 3D images (second row) with different percentages of randomly located occlusion from 0% to 50%.

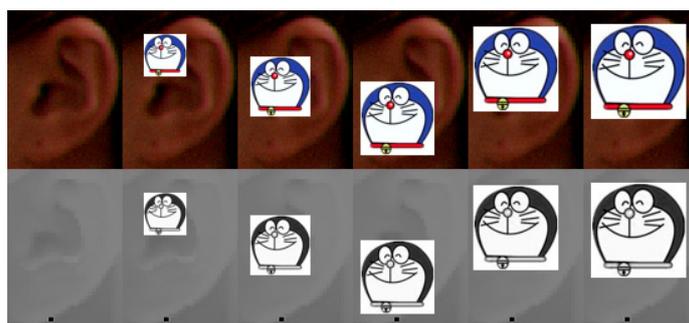


Figure 2. Different percentages of occlusion in the probe images.

To evaluate the performance of the presented ear detection method, as done in [14], we used the indexes precision, recall, accuracy, and F1 score. The experimental results obtained are summarized in Table 1.

Table 1. Detection results of different occlusion rates.

Number of Images in Dataset	Occlusion Rate (%)	Precision (%)	Recall (%)	Accuracy (%)	F1 Score (%)
1800	0	100	100	100	100
1800	10	100	100	100	100
1800	20	100	100	100	100
1800	30	99.9	100	99.9	99.9
1800	40	99.9	99.6	99.6	99.7
1800	50	99.5	98.9	98.6	99.2

With the 3D ear detected, the 3D data contain depth discontinuities such as spikes and holes. To remove these spikes from the depth scans, we applied a median filter, and, to fill the holes, we used cubic interpolation. To reduce noise and smoothen the 3D data, a Gaussian filter was used.

4. Local Feature Representation

The bounding boxes represent the outputs of ear detection, and the regions of interest containing the ears can be cropped from them and passed on to the feature extraction stage.

4.1. Definition of Features

Curvature is an intrinsic property of any surface, and it has been shown to play an important role in 3D ear recognition [11,33]. The two functions of curvature introduced by Koenderink [52], “shape index” and “curvedness,” were found to convey far more information than the usual mean and Gaussian curvatures. The former is a measure of “which” shape captures the intuitive notion of the “local shape”; the latter is a measure of “how much” the shape is a positive number that specifies the amount of curvature. On the basis of the two local principal curvatures, k_{max} and k_{min} , computed numerically at each point p of the 3D surfaces, the shape index (S_I) and curvedness (C_v) [53] were computed.

$$S_I(p) = \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{k_{max}(p) + k_{min}(p)}{k_{max}(p) - k_{min}(p)}\right) \quad (3)$$

$$C_v(p) = \sqrt{\frac{k_{min}^2(p) + k_{max}^2(p)}{2}} \quad (4)$$

4.2. Keypoint Selection

The aim of this step is to select a repeatable and salient set of potential keypoints. Inspired by [31], Zhou et al. [11] proposed a novel keypoint detection technique to choose points with higher curvedness values than other points within a small neighborhood; those points contain salient surface information and are highly distinctive. A square window with dimensions 1 mm × 1 mm is used to scan the segmented ear region, and the point with the highest curvedness value within each small neighborhood is marked as a keypoint. The local surface data around each keypoint are cropped using a sphere centered at the keypoint. To eliminate the insignificant keypoints chosen in the former procedure, the data of the cropped neighborhood surface of the selected keypoints are examined and, the selected keypoints within the neighborhood that contain boundary points are eliminated. To further reject the less discriminative keypoints, PCA [54] is applied to the cropped neighborhood surface data, and the eigenvalues are calculated to judge the discrimination power associated with each keypoint. If the largest and the smallest eigenvalues of these keypoints satisfy the predefined thresholds in [11], they are retained. We adopted this keypoint selection approach.

4.3. Local Feature Representation

The SPHIS representation of a 3D surface can be used directly for a local patch feature representation [11]. Let p and S be the keypoint and surface patch cut by a sphere with a radius

of 14 mm centered at p . S_1, S_2, S_3, S_4 correspond to four surface patches having distances between each point contained in it and p of less than 3.5 mm, from 3.5 to 7 mm, from 7 to 10.5 mm, and from 10.5 to 14 mm, respectively.

For patch S_k , p_j represents a point contained in it. $S_I(p)$ and $C_V(p)$ correspond to the shape index and curvedness of point p_j and are calculated by [53]. The set $\{x_{k,i} | i = 1, 2, \dots, 16\}$ contains the centers of the bins. The weight $h(x_{k,i})$ of the k^{th} bin is calculated by

$$h(x_{k,i}) = \sum_{j=1}^n Co(p_j, x_{k,i}) Co(p_j, x_{k,i}) = \begin{cases} C_V(p_j) \left(1 - \frac{S_I(p_j) - x_i}{0.0625}\right) & x_i \leq S_I(p_j) < \min(x_{i+1}, 1 + \zeta) \\ C_V(p_j) \left(\frac{S_I(p_j) - x_i}{0.0625}\right) & \max(0, x_{i-1}) \leq S_I(p_j) < x_i \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where $\zeta \rightarrow 0 \cap \zeta \in R^+$.

The HIS descriptor of S_k is $\frac{(h(x_{k,1}), h(x_{k,2}), \dots, h(x_{k,16}))}{\sum_{i=1}^{16} h(x_{k,i})}$, and the SPHIS descriptor of S is formed by concatenating the HIS descriptors of $S_1, S_2, S_3, S_4, S_I(p)$, and $C_V(p)$.

4.4. Two-Step NGT-Based Method for the Local Surface Matching Engine

This section describes a modified NGT-based matching method for quickly finding multiple correspondence keypoints that satisfy a rigid constraint and theoretically justifies its performance. The method allows a user to accurately align one gallery range image with one probe range image containing a portion of outliers and obtain as many keypoint correspondences as possible for every probe-gallery pair. We developed our own variant of this algorithm. The steps in the Algorithm 1 are as follows:

Algorithm 1 Two-step NGT-based method for two images alignment

- 1: Build the initial set of strategies S
 - 2: Generate a set of keypoint correspondences maintaining a rigid constraint
-

In step 1, to generate set S , we first apply a proposed breed NGT-based matching method to find a core set of strategies S' by using the local information that is unaffected by occlusion. Then, we expand set S' into set S by evaluating the geometrical consistency of all of the keypoints from every probe-gallery pair.

For a given probe range image P and gallery range image G , we define an alternative set of strategies S_1 as the input set of the NGT-based matching method:

$$S_1 = \{(a, b) \in P \times G | b \in dn_k(a) \cap \|SPHIS(a) - SPHIS(b)\|_2 < 0.1\} \quad (6)$$

where a is a probe keypoint, b is a gallery keypoint, and $dn_k(a)$ is the set of k gallery ear keypoints with the nearest SPHIS descriptor with respect to the SPHIS descriptor of a . In a noncooperative game, the value of k is specified as 5 [10]. $SPHIS(\bullet)$ is the SPHIS descriptor of the keypoint \bullet . $\|\cdot\|_2$ represents the L_2 distance between the feature descriptors.

Then, to select a larger set of keypoint correspondences satisfying the rigid constraint, we define an earnings function e :

$$e((a_1, b_1), (a_2, b_2)) = \frac{\min(\|a_1 - a_2\|_2, \|b_1 - b_2\|_2)}{\max(\|a_1 - a_2\|_2, \|b_1 - b_2\|_2)} \quad (7)$$

This function takes pairs of strategies $(a_1, b_1), (a_2, b_2) \in P \times G$ and allots an earnings e in accordance with the distances exhibited by the strategy pair. In addition, to avoid possible many-to-many matches, a hard constraint is imposed by setting the compatibility between two

strategies that share the same probe or gallery keypoint to zero [9]. Furthermore, for the local surface matching engine proposed in [11], to remove correspondences that do not maintain a rigid constraint, for a correspondence composed of the centroids of the probe and gallery keypoints and another correspondence composed of a probe and gallery keypoint, after probe and gallery model alignment, if the absolute value of the difference between two distance values from the distance between two keypoints from the probe model and the distance between two keypoints from the gallery model contained in two correspondence is larger than 1.5 mm, the two correspondences cannot coexist in the final local surface matching result. Inspired by this method, in order to satisfy the rigid constraint, for an arbitrary strategy pair, after calculating the distance between two keypoints from the probe model and the distance between two keypoints from the gallery model, if the absolute value of the difference between two distance values is greater than the threshold of 1.5 mm, the earnings of the strategy pair is set to zero. Thus, the final earnings is fine-tuned.

$$C = \begin{cases} e((a_1, b_1), (a_2, b_2)) & \text{if } a_1 \neq a_2 \text{ and } b_1 \neq b_2 \text{ and } \|a_1 - a_2\|_2 - \|b_1 - b_2\|_2 \leq 1.5mm \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Subsequently, on the basis of a state-of-the-art noncooperative game, the candidate set S_1 and the earnings function are used to generate the earnings matrix. To quickly evolve to a Nash equilibrium, we use the infection–immunization dynamics introduced by Rota Bulò and Bomze [55], which have $O(N)$ complexity for each step. To the best of our knowledge, this is the most computationally efficient evolutionary selection process. When a stable state is reached, the occupied share of all strategies is concatenated to vector X . The maximum element in X is denoted as $\max(X)$. The final solution at the equilibrium is composed of these strategies, where each of these strategies occupies more than a $\frac{\max(X)}{2}$ share. All of these strategies construct a set S' .

When set S' is obtained, the next task is to generate set S that can include as many ground-truth keypoint correspondences as possible. To achieve this goal, we first find all of the keypoints from set S' that belong to probe image P . Then, the remaining keypoints from probe image P are used to construct a set of keypoints P_k . For a given keypoint kp from P_k and all of the keypoints from a gallery range image G , we define a set of strategies S_{P_k} :

$$S_{P_k} = \{(kp, b) \in P \times G \mid \|SPHIS(kp) - SPHIS(b)\|_2 < \Gamma, kp \in P_k, b \in G\} \quad (9)$$

To obtain the value of the threshold, Γ , we first align the corresponding probe and gallery range images, find all of the outlier correspondence voxels (the details from step 1 to step 3 of the online stage are presented in Algorithm 2), and then define Γ :

$$\Gamma = \frac{NVOC}{NV} \quad (10)$$

where $NVOC$ and NV are, respectively, the number of voxels associated with the outlier correspondences and the number of voxels (the details of step 1 of the online stage are presented in Algorithm 2) in the local spherical region with a radius of 14 mm centered at the keypoint kp . In this study, if the center point of the voxel belonged to the local region, then the corresponding voxel also belonged to the local region. Furthermore, we found that the ratio of the number of voxels associated with the outlier correspondences to the number of voxels in the local region centered at every keypoint was usually less than 50% by observing the ear range images with partial occlusion (as shown in Figure 2 and Section 7.2). Thus, when the value of Γ is greater than 0.5, the keypoint kp does not match any of the keypoints from the corresponding gallery range image G .

For the i^{th} strategy from set S_{P_k} , we calculate the earnings between it and each of the strategies from S' by using Equation (8) and concatenate all of the earnings to a set. The number of elements in this set that are larger than zero is represented as ne_i . In consideration of the arbitrariness of i , a set is

formed by concatenating $ne_1, ne_2, \dots, ne_{nS}$, and the maximum value in this set is obtained, where nS is the number of elements in set S_{Pk} . In set S_{Pk} , all of the strategies associated with the maximum value are extracted and concatenated to set S . In consideration of the arbitrariness of kp , for each of these keypoints from Pk , we can extract some corresponding strategies and concatenate these strategies to set S . Finally, set S' is concatenated to set S .

In step 2, we propose another breed NGT-based matching method for finding a set of strategies by using the initial set S . When the initial set S is input, we construct the earnings function, generate the earnings matrix, evolve to a Nash equilibrium, and obtain a final set of strategies maintaining a rigid constraint following these methods described in step 1. In a stable state, the keypoint correspondences associated with all of the strategies retained as correct matches are used to calculate the transformation approximation and align the corresponding probe and gallery range images by using a least-squares fitting technique [56]. Following application of the proposed NGT-based matching method, the local surface matching engine outputs the number of keypoint correspondences for every probe-gallery pair as the similarity score. Figure 3 demonstrates an example of recovering the keypoint correspondences from a pair of gallery and probe ear models.

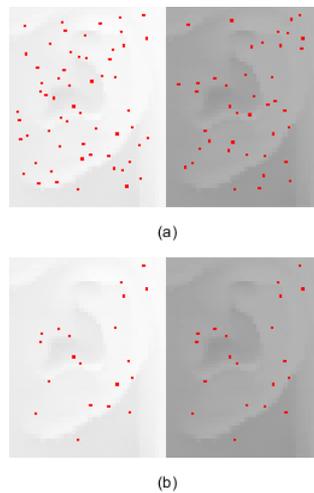


Figure 3. Example of finding the feature correspondences for two ears from the same subject. (a) Keypoint detected on ears. (b) True feature correspondences recovered by local matching engine. The red dots in (a) indicate the keypoints to be acquired at the keypoint selection stage. The red dots in (b) represent the reserved keypoints by the local matching engine.

We justify the effectiveness of the proposed two-step NGT algorithm in this section based on the theory described in [57]. In the first step of the NGT algorithm, according to Theorem 2 described in [57], and the definition of the alternative set of strategies S_1 , $S_1 \subseteq M_a \times D$ is considered to be a set of matching strategies over M_a and D with $(m, Tm) \in S_1$ for all $m \in M_a \cap M_b$, where $M_a \subseteq M$, $M \subseteq P$ is a set of keypoints contained in probe image P . $D = TM_b$ is a rigid transformation of $M_b \subseteq M$ such that $|M_a \cap M_b| \geq 3$, the keypoints are not in the overlap; that is, the keypoints in $E_a = M_a(M_a \cap M_b)$ and $E_b = M_b(M_a \cap M_b)$, are sufficiently far away such that for every $s \in S_1$, $s = (m, Tm)$ with $m \in M_a \cap M_b$ and every $q \in S_1$, $q = (m_a, Tm_b)$ with $m_a \in E_a$ and $m_b \in E_b$, we have $\pi(q, s) < \frac{|M_a \cap M_b| - 1}{|M_a \cap M_b|}$, then, the vector $\hat{x} \in \Delta^{|S_1|}$, defined as

$$\hat{x}_i = \begin{cases} 1/|M| & \text{if } s_i = (m, Tm) \text{ for some } m \in M_a \cap M_b \\ 0 & \text{otherwise} \end{cases}$$

is an evolutionary stable strategy (ESS).

When the ESS is reached, the output of the first step of the NGT algorithm is a set of strategies S' that correspond to ESS. For every $s \in S'$, the corresponding $\hat{x}_i > 0$; thus, $S' \subseteq (M_a \cap M_b) \times T(M_a \cap M_b)$ with $(m, Tm) \in S'$ for all $m \in (M_a \cap M_b)$.

In the second step of the NGT algorithm, according to the definition of strategies S , $S \subseteq M_a^s \times D^s$ is considered to be a set of matching strategies over M_a^s and D^s , with $(m, T^s m) \in S$ for all $m \in M_a^s \cap M_b^s$, where $M_a^s \subseteq M^s$, $M^s \subseteq P$ is a set of keypoints contained in P . $D^s = T^s M_b^s$ is a rigid transformation of $M_b^s \subseteq M^s$ such that $|M_a^s \cap M_b^s| \geq 3$, the keypoints are not in the overlap; that is, the keypoints in $E_a^s \subseteq M_a^s (M_a^s \cap M_b^s)$ and $E_b^s \subseteq M_b^s (M_a^s \cap M_b^s)$, are sufficiently far away such that for every $s \in S$, $s = (m, T^s m)$ with $m \in M_a^s \cap M_b^s$ and every $q \in S$, $q = (m_a^s, T^s m_b^s)$ with $m_a^s \in E_a^s$ and $m_b^s \in E_b^s$, we have $\pi(q, s) < \frac{|M_a^s \cap M_b^s| - 1}{|M_a^s \cap M_b^s|}$; then, the vector $\hat{x}^s \in \Delta^{|S|}$, defined as

$$\hat{x}_i^s = \begin{cases} 1/|M^s| & \text{if } s_i^s = (m, T^s m) \text{ for some } m \in M_a^s \cap M_b^s \\ 0 & \text{otherwise} \end{cases}$$

is an ESS.

When the ESS is reached, the output of the second step of the NGT algorithm is a set of strategies S^s corresponding to ESS. For every $s \in S^s$, the corresponding $\hat{x}_i^s > 0$; thus, $S^s \subseteq (M_a^s \cap M_b^s) \times T(M_a^s \cap M_b^s)$ with $(m, Tm) \in S^s$ for all $m \in (M_a^s \cap M_b^s)$.

We verify that the two-step NGT algorithm outperforms the traditional one-step NGT algorithm in the following two cases: $T = T^s$ and $T \neq T^s$. When $T = T^s$, $S' \subseteq S$ and $S' \subseteq (M_a \cap M_b) \times T(M_a \cap M_b)$ and $S \subseteq M_a^s \times D^s$, so $(M_a \cap M_b) \subseteq M_a^s$ and $T(M_a \cap M_b) \subseteq D^s$. $D^s = T^s M_b^s$, as a result $(M_a \cap M_b) \subseteq M_b^s$; thus, $(M_a \cap M_b) \subseteq (M_a^s \cap M_b^s)$. According to the definition of S' and S^s , we have $S' \subseteq S^s$; thus, $|S'| \leq |S^s|$. For every P-G pair, a higher local similarity score is obtained in the second step of the NGT algorithm than that in the first step of the NGT algorithm, $S^s S' \subseteq ((M_a^s \cap M_b^s)(M_a \cap M_b)) \times (T(M_a^s \cap M_b^s)T(M_a \cap M_b))$. Let OA be the occluded region contained in P , and AO be the ear surface region contained in P , where the local neighborhood of the keypoints contained in it is affected by occlusion, for every keypoint $kp \in (M_a^s \cap M_b^s)(M_a \cap M_b)$, kp satisfies Equation (9) and $\Gamma \leq 0.5$. $kp \in AO$ because $T(kp) \notin G$ when $T(kp) \in OA$; thus, $(M_a^s \cap M_b^s)(M_a \cap M_b) \subseteq AO$. Let $num1 = |(M_{a1}^s \cap M_{b1}^s)(M_{a1} \cap M_{b1})|$ be the difference between the local similarity scores of the one-step NGT algorithm and the two-step NGT algorithm with the same identity of a probe image P and a gallery image G_1 . Let $num2 = |(M_{a2}^s \cap M_{b2}^s)(M_{a2} \cap M_{b2})|$ be the difference between the local similarity scores of the one-step NGT algorithm and the two-step NGT algorithm with the same identity of a probe image P and a gallery image G_2 . Then, $num1 \geq num2$ because the keypoints in the local region of P are easily matched to the keypoints in the corresponding region of a gallery image with the same identity, but not with the keypoints in the region of a gallery image with different identity. For P and G_1 , let LS_{11} be the local similarity score for the first step of the NGT algorithm and LS_{12} be the local similarity score for the second step of the NGT algorithm. For P and G_2 , let LS_{21} be the local similarity score for the first step of the NGT algorithm and LS_{22} be the local similarity score for the second step of the NGT algorithm. $LS_{12} = LS_{11} + num1$, $LS_{22} = LS_{21} + num2$. As the local similarity score with the same identity is greater than that of two images with different identities, $LS_{11} \geq LS_{21}$. Thus, $LS_{12} \geq LS_{22}$. For every P and G_1 with the same identity and P and G_2 with different identities, the local similarity score of P and G_1 is greater than that of P and G_2 .

When $T \neq T^s$, we have $|S^s| \leq |((M_a^s \cap M_b^s)(M_a \cap M_b))| + 2$.

Proof. Hypothesis $|S^s| > |((M_a^s \cap M_b^s)(M_a \cap M_b))| + 2$, at least \exists keypoints $p_i \in M_a \cap M_b, T(p_i) = T^s(p_i)$, $i = 1, 2, 3$. Because p_1, p_2, p_3 are not collinear in general, the three pairs of corresponding points of a collinear only determine a rigid transformation, we have $T = T^s$ in general, which is in contradiction with $T \neq T^s$. Thus, $|S^s| \leq |((M_a^s \cap M_b^s)(M_a \cap M_b))| + 2$. \square

As $(M_a^s \cap M_b^s)(M_a \cap M_b) \subseteq AO$, $(M_a \cap M_b) \subseteq ESE$, ESE is the ear surface region contained in P , where the local neighborhood of keypoints contained in it is not affected by occlusion. $|S^s| \leq |((M_a^s \cap M_b^s)(M_a \cap M_b))| + 2 \ll |M_a \cap M_b|$, ESE is difficult to match with G . Thus, we can only get smaller local similarity scores, P and G correspond to different identity. Ear recognition results will not be affected.

The experiments in the benchmark dataset show that our NGT-based matching algorithm outperforms the traditional NGT-based matching algorithm. The reason is that our method can estimate the occlusion rate of a local surface patch of every keypoint by first performing NGT-based matching. Then, this information is used to associate many keypoints at which the local surface patch is seriously affected by occlusion with their true correspondence. In contrast, traditional NGT-based matching algorithms have difficulty associating these keypoints with their true correspondence.

5. Holistic Feature Extraction

5.1. Preprocessing

The preceding section described the method by which correspondences are established between a probe-gallery pair. The probe model is then registered onto the gallery model by using the transformation obtained by the local matching stage for each point on the probe model. When the amount of established correspondences is below three, we adopt the ear pose normalization technique described in [11] for model registration.

5.2. Holistic Representation

The holistic representation used in this step is a voxelization of the surface. The motivation behind using such a feature is that it has many advantages, i.e., a higher efficiency, noise suppression capacity, a compact representation of the 3D surface, and the ability to encode the surface features of the subject [11]. Experimental results for the University of Notre Dame J2 database have demonstrated the feasibility and effectiveness of surface-voxelization-strategy-based ear recognition methods.

The representation applied in this step is binary voxelization. This representation effectively encodes the presence of a point inside a voxel. A voxel that has a point enclosed within it is distributed a value of "1"; otherwise, it is distributed a value of "0." Zhou et al. [11] described a voxelization process using this representation. We adopt this voxelization process to perform holistic representation.

5.3. Holistic Feature Matching Engine

Zhou et al. [11] proposed a sophisticated voxelized surface matching technique to calculate the holistic similarity between two subjects. In the offline enrollment step, for a designated gallery model, a voxel grid is generated from the bounding volume enclosing the model. The gallery model is then voxelized, and this representation is registered on the gallery. In the online step, the transformation used to register a probe-gallery model pair in the local matching step is applied to the bounding volume of the probe model. The joint spatial expansion of the registered probe and gallery model bounding volumes is then calculated. The voxel grid applied to voxelize the gallery model is extended to enclose both bounding volumes, and the probe model is voxelized by this extended voxel grid. Furthermore, the voxelization representation of the gallery model is zero-padded. The vectors \overline{V}_p and \overline{V}_g corresponding to the probe and gallery models, respectively, are produced by voxelizing both models and vectorizing the voxelizations. They calculate the similarity between these vectors by using the cosine similarity measure given by

$$S(p, g) = \frac{\overline{V}_p \cdot \overline{V}_g}{\|\overline{V}_p\| \cdot \|\overline{V}_g\|} \quad (11)$$

Zhou et al. proposed a voxelized surface matching technique that has achieved satisfactory performance on the benchmark dataset. However, when a partial ear surface is occluded, an outlier from the occlusion region is usually farther from its associated point lying on the ear surface (as shown in Figure 4). Although a sharp increase in the size of the voxel can reduce the impact of the outlier [11], the experimental results in [11] show that a low recognition rate is only obtained in that case. To decrease the impact of outliers, some holistic-registration-based ear measurement methods [8] remove abnormal points using the distance threshold approach. Before the distance threshold is obtained, these methods need to build a correspondence between the surface points of the two models. There are several methods for finding a correspondence in the field of 3D subject recognition, e.g., the closest-point matching strategy and the “projection” matching strategy. Related studies have shown that the “projection” matching strategy achieves better performance against outliers [58]. However, the “projection” matching strategy has been found to potentially generate a large number of incorrect pairings of a voxel on an uneven 3D ear surface (as shown in Figure 5, the “many-to-one” phenomenon is common, resulting in partial outliers that are not allowed to be deleted in subsequent steps [58]). Therefore, the traditional holistic ear surface matching technique is not suitable for solving the holistic matching problem of a voxelized ear surface containing partial occlusion.

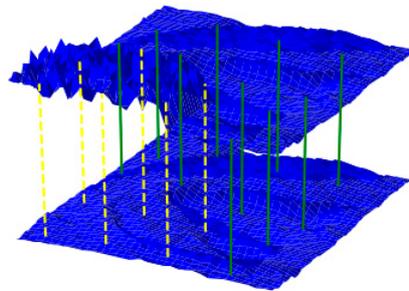


Figure 4. Probe ear range image with partial occlusion and gallery ear range image with good image quality registration. The yellow dashed lines connect the occluded points in the probe model with their associated closest points in the gallery model. The green lines connect the surface points in the probe model with their associated closest points in the gallery.

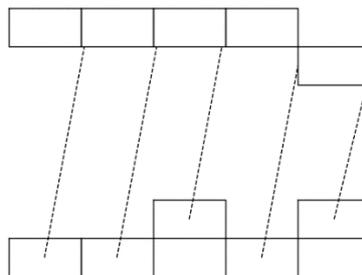


Figure 5. In the case of an uneven voxelized surface, the “projection” matching strategy can generate a large number of incorrect pairings. The dashed lines connect the surface voxels in the voxelized probe model to their associated surface voxels in the voxelized gallery model.

To use holistic features to recognize a voxelized ear, we propose a variant of the breed voxelized surface matching technique. The proposed technique combines the advantages of the surface voxelization method and the idea of using the distance threshold to remove an occluded region. Moreover, we propose a “normal projection” matching strategy to build the correspondence between the surface points of the two models. The steps in the proposed variant of the breed voxelized surface matching technique are as follows:

Algorithm 2 Variant of breed voxelized surface matching

In the offline enrollment stage:

1: Denote the normal direction \overline{m}_r of the gallery model

$$\overline{m}_r = \frac{1}{n_r} \sum_{i=1}^{n_r} \overline{p}_i \quad (12)$$

where \overline{p}_i represents the normal of the i^{th} surface point, and n_r is the total number of surface points contained in the gallery model.

2: Construct a bounding box enclosing the gallery model, where the bottom of the bounding box is perpendicular to \overline{m}_r .

3: Voxelize the gallery model using a voxel grid constructed from the bounding box enclosing the gallery model [11].

In the online stage:

1: Calculate the joint spatial extent of the registered probe-model and gallery-model bounding boxes and the voxel grid enclosing the bounding box [11].

2: Choose the corresponding voxels between the registered probe and gallery models within the bounding boxes by proposing a “normal projection” matching strategy

2.1: Determine the set of Gv gallery voxels $\{v_{g_i} | i = 1, 2, \dots, Gv\}$ and a set of Gv centers of gallery voxels $\{c_{g_i} | i = 1, 2, \dots, Gv\}$, where c_{g_i} is the center of v_{g_i} .

2.2: For $i = 1, 2, \dots, Gv$, do the following.

2.3: Draw a straight line l_i through c_{g_i} ; its direction is the same as \overline{m}_r .

2.4: Find the voxel v_{p_i} that intersects l_i in the corresponding probe model P . Assign v_{g_i} and v_{p_i} as a pair.

2.5: End for.

3: Remove the abnormal pairs of voxels.

4: Perform voxelized subject vectorization; the vectors \overline{V}_p and \overline{V}_g correspond to the probe and gallery models

5: Calculate the registration error of the two models using Equation (11).

For step 2 of the online stage, an example is given in Figure 6. For the two straight lines passing through the two centers of two voxels along the direction of the normal \overline{m}_r (the direction is perpendicular to the bottom of the bounding box), the distance between them is not less than the size of the voxel. Thus, any two voxels that intersect any two straight lines are completely different. Incorrect pairings are difficult to obtain.

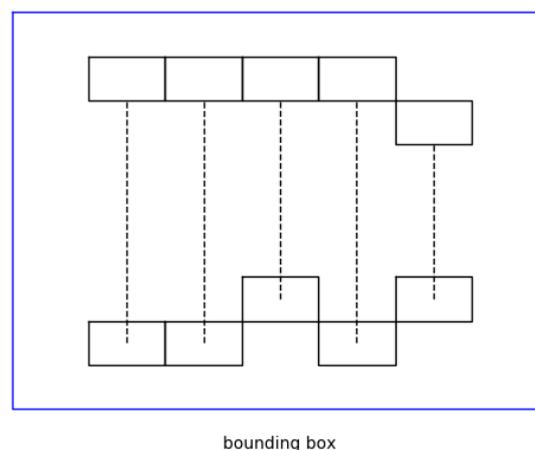


Figure 6. Projection of the bounding box containing the voxelized surface of the range image, where the normal of the range image is perpendicular to the bottom of the bounding box.

In step 3 of the online stage, we first compute the distance (denoted as the distance between the center points of the two associated voxels) per pair of voxels. Then, the parts of the voxel pairs that are too far apart are discarded using the tolerance distance tol . Furthermore, all of the retained distance

values are sorted, and the highest th percent are deleted. The matrices V_g and V_p are then updated. To determine the optimal distance tol and th , the values of the distance tol in the range 4–14 mm in intervals of 2 mm and the values of percent th (1, 5, 10, 15, 19, 20, and 30) were tested; the values $tol = 10$ mm and $th = 10$ gave the best performance.

In steps 4 and 5 of the online stage, we use the relevant step in the voxelization method proposed by Zhou et al. [11]. In our experiment, for comparison with SOA ear recognition methods based on voxelization [11], only cubed voxels were considered, and the size of a voxel was specified as 1 mm.

Our voxelization method differs from that of Zhou et al. in that we have added a distance threshold to the basic voxelization method to avoid matching any voxel v_{p_i} of one surface to a remote part of another surface that is likely to not correspond to v_{p_i} . Such a voxel v_{p_i} from the surface of the probe range image P might be from a portion of the scanned object that was not captured in the gallery range image G ; thus, no pairing should be made to any point on G . We have found that robust occlusion identification results when the distance threshold is set appropriately.

6. Fusion

Two types of match score measurement components result in unattached similarity matrices S_i , each of size $s(P) \times s(G)$, where $i \in \{1, 2\}$ denotes the measurement engine, and $s(P)$ and $s(G)$ indicate the numbers of probes and gallery range images, respectively. We use the transform-based decision-making integration technique proposed in [11] to fuse these two measurement scores. First, they are transformed into the scope of $[0, 1]$ by a double sigmoid normalization scheme [59]; then, the weighted sum of the normalized scores is used to compute the last measurement score.

7. Experimental Results and Discussion

In this section, we report on extensive experiments conducted to evaluate the proposed method in different scenarios. These experiments consisted of ear recognition with natural occlusion and random occlusion. All experiments were carried out using a 3 GHz PC with 4 GB of RAM. As the data fusion technique requires training six parameters, we needed to train the associated parameters beforehand.

7.1. Training of the Data Fusion Parameters

To train the six data fusion parameters—namely, $\alpha_1^l, \alpha_2^l, \tau^l, \alpha_1^h, \alpha_2^h$, and τ^h —Zhou et al. proposed a parametric training method [11]. In their method, they first initialize the two τ parameters to the 10 th percentile of the genuine match scores of their respective modalities, which are bounded to the limits of the $[0, 20th]$ percentile of the genuine match scores of their respective modalities. Then, α_1 and α_2 are initialized to half the τ value of their respective modalities. Finally, they obtain these parameters by minimizing the EER with respect to the parameters. Constrained optimization is then performed using the GlobalSearch framework provided by MATLAB. We adopted this parametric training method to train the associated parameters.

The experimental results for a benchmark dataset indicated that the performance of local surface matching and the holistic matching method is easily affected by the percentage of occlusion. To make the obtained data fusion parameters applicable to ear data with various percentages of occlusion, we constructed a training set comprising 300 images of 300 distinct subjects from Single4 and constructed a test set generated using the remaining subjects from Single4. This set of the remaining subjects from Single4 (which comprised 115 images) was labelled Single5. The process was repeated four times, and in each iteration a distinct portion of Single4 was designated as the training set. The results presented in the previous section were obtained by computing the mean performance across the four folds of the cross validation.

The process of constructing the test set is as follows. In the first step, two candidate test sets are constructed. Then, a test set comprising six ear data per subject is constructed from these candidate test sets.

To construct the first candidate test set, we used scanned ear data with occlusion due to hair and earrings. We first normalized 115 images from Single5 and recorded Γ_i , which is the transformation relationship between the i^{th} probe ear datum c_i and the corresponding normalized datum s_i :

$$c_i = \Gamma_i(s_i) \quad (13)$$

where $i = 1, 2, \dots, 115$. Subsequently, 18 scanned images with occlusion due to hair and 21 scanned images with occlusion due to earrings were selected and normalized, and 39 occlusion blocks of different shapes extracted. For each normalized ear datum s_i , we randomly selected a type of occlusion block to add to the data and simulated the candidate testing data by using a Γ_i transformation for the corresponding normalized ear data.

As shown in Section 7.3, each subject contained in Single5 generated six images with six distinct occlusion percentages. As a result, 690 images were generated by the 115 distinct subjects contained in Single5. We used these images to construct the second candidate test set.

To generate the test set by using two candidate test sets, we first established six categories with reference values of 0%, 10%, 20%, 30%, 40%, and 50%. Each ear datum from the two candidate test sets were then assigned a category tag with a reference value closest to its percentage of occlusion. Then, each of the seven ear data of each subject was assigned a class tag. For each subject, if the class tag of its ear datum differed from another ear datum, then the ear datum had a probability of one; if the class tag of two ear data were the same, then the two ear data had a probability of 1/2. Finally, we randomly selected six from seven ear data of each subject according to the specified probability to construct a test set.

7.2. Ear Recognition with Natural Occlusion

Our experiments were performed using the UND-J2 database, which comprises 1800 3D range images and co-registered color images of 415 people, with data acquired using a Minolta Vivid 910 camera. Among the images, the range images of the ears of 35 people were found to be partially covered by minor amounts of hair, and those of 42 people were partially covered by earrings. Figure 7a shows two examples of occlusion by minor amounts of hair. Figure 7b shows two examples of occlusion by earrings. The first row of Figure 7 shows 2D images of ears, and the second row shows the 3D depth images.

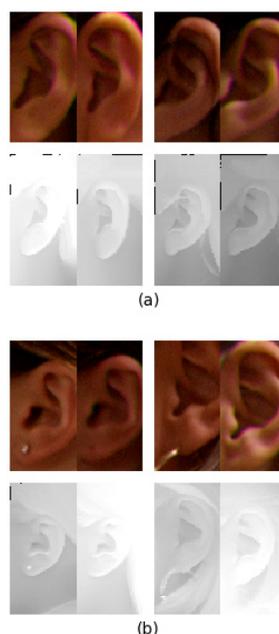


Figure 7. Examples of occlusion by (a) small amount of hair and (b) earrings.

To evaluate the robustness against natural occlusion, we conducted two experiments. For the first experiment, 35 range images occluded by hair in the database were used to constitute a probe set H . Then, we used 35 range images with good image quality to form a gallery set $Single1$, where the identity of the i^{th} range image from $Single1$ is the same as the identity of the i^{th} range image in the related probe set. For the second experiment, we followed the method below to construct a probe set and gallery set, replacing hair with earrings, replacing 35 with 42, replacing H with E , and replacing $Single1$ with $Single2$.

The results of these experiments are presented in Table 2. The second and third columns of Table 2 respectively present the rank-1 recognition rate for occlusion due to small amounts of hair and that for occlusion due to earrings. The second, third, and fourth columns of Table 2 are, respectively, the identification performance of our local surface matching (SPHIS + two-step NGT), the identification performance of the measurement based on our holistic matching (the variant of breed surface voxelization), and the identification performance of our fusion method (our proposed method).

Table 2. Rank-1 recognition rates for the identification experiments of H vs. Single1 and E vs. Single2.

	SPHIS + Two-Step NGT	A Variant of Breed Surface Voxelization	Our Proposed Method
H vs. Single1	97.1%	94.3%	97.1%
E vs. Single2	97.6%	97.6%	97.6%

To compare the discrimination potential of our proposed method with the basic NGT-based recognition method, we compared our proposed method and the SPHIS + two-step NGT-based recognition method to a baseline: SPHIS + the basic NGT-based recognition method [9]. The experiments were carried out independently for the H and $Single1$ datasets and for the E and $Single2$ datasets. Tables 3 and 4 summarize the rank-1 recognition rates. We can see that the SPHIS + two-step NGT-based recognition method and our proposed method attain higher rank-1 recognition rates than the SPHIS + NGT-based recognition method for the two experiments. Our proposed method and the SPHIS + two-step NGT-based recognition method slightly outperform the baseline when applied to 3D ear recognition in the natural occlusion scenario.

Table 3. Comparison of the recognition performance of two-step NGT and basic NGT with occlusion due to hair.

Method	Gallery	Probe	Rank 1
Our proposed method	Single1	H	97.1%
SPHIS + two-step NGT	Single1	H	97.1%
SPHIS + NGT [9]	Single1	H	94.3%

Table 4. Comparison of the recognition performance of two-step NGT and basic NGT with occlusion due to earrings.

Method	Gallery	Probe	Rank 1
Our proposed method	Single2	E	97.6%
SPHIS + two-step NGT	Single2	E	97.6%
SPHIS + NGT [9]	Single2	E	95.2%

7.3. Ear Recognition with Random Occlusion

In real-world applications, captured ear images are likely to be occluded by other occluding objects. In this set of experiments, we chose one image with good image quality for each person as a gallery sample and one image for each subject not included in the gallery as a probe sample.

The gallery and probe set were respectively labelled Single3 and Single4. We followed the method described in Section 3 to simulate different intrusive occluding objects in every probe sample.

The ear recognition method proposed in this paper is a decision-layer fusion method. As described in Sections 4 and 5, a measurement based on SPHIS + two-step NGT and that based on a variant of breed surface voxelization were used in the identification phase. The final identification result was obtained via a weighted fusion algorithm. Figure 8 shows the rank-1 recognition rate as a function of the number of candidates.

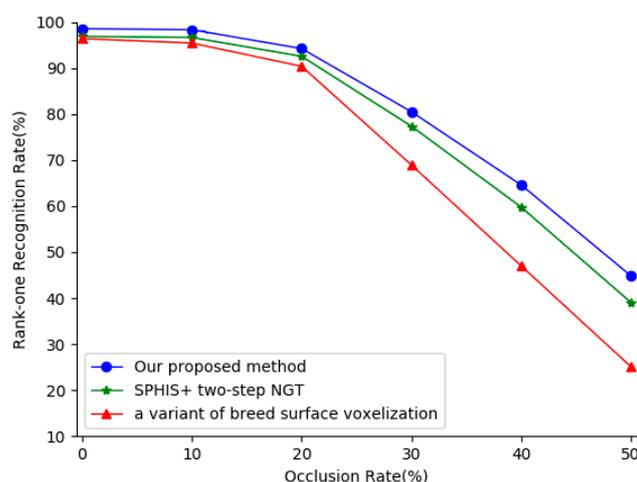


Figure 8. Recognition rates of our proposed method, SPHIS + two-step NGT, and a variant of the surface voxelization method for various random occlusion rates.

We evaluated the identification performance of three SOA systems, described in [20], [11], and [8], under partial occlusion. A comparison of these methods is provided in Table 5. The recognition rates of our method remain relatively steady. This is primarily because more keypoints with a local neighborhood affected by occlusion contained in the probe image are successfully matched with the associated keypoints of the gallery image in the local matching component and the location of the ear pit is not required prior to matching.

Table 5. Comparison of the recognition performance for different percentages of random occlusion.

Technique	Percentage of Occlusion					
	0%	10%	20%	30%	40%	50%
Yan and Bowyer [20]	97.8%	73.5%	45.1%	25.8%	8.9%	1.4%
Zhang et al. [8]	98.6%	74.2%	46.0%	26.7%	9.9%	2.4%
Zhou et al. [11]	98.3%	95.9%	90.1%	65.1%	40.5%	18.8%
Our proposed method	98.8%	98.3%	94.2%	80.5%	64.6%	44.8%

In Table 6, we provide a run-time efficiency comparison of the three SOA 3D ear biometric systems evaluated on the same database. The researchers in [20] and [8] proposed systems implementing ICP-based algorithms for shape recognition, which are time consuming, especially when applied to high-resolution dense samples, as in the SOA systems. In contrast, the researchers in [33] and this paper proposed approaches that employ a sparse set of features that can substantially reduce the number of vertices considered when registering samples.

Table 6. Performance comparison to other 3D ear recognition systems.

Method	Run Time	Ear Detection
Yan and Bowyer [20]	5–8 s	Automatic
Zhang et al. [8]	2.72 s	Automatic
Zhou et al. [11]	0.02 s	Automatic
Our proposed method	0.019 s	Automatic

Following the method described in the last paragraph of Section 7.2, we compared our method and the SPHIS + two-step NGT-based recognition method to the SPHIS + basic NGT-based recognition method with the probe and gallery sets generated in this section. Figure 9 shows the rank-1 recognition rates for different occlusion rates. The improvement for the ear database is even more significant, as our proposed method and SPHIS + two-step NGT-based recognition method are respectively greater than 10% and 5% higher than the basic NGT-based recognition method for each sub-experiment on average.

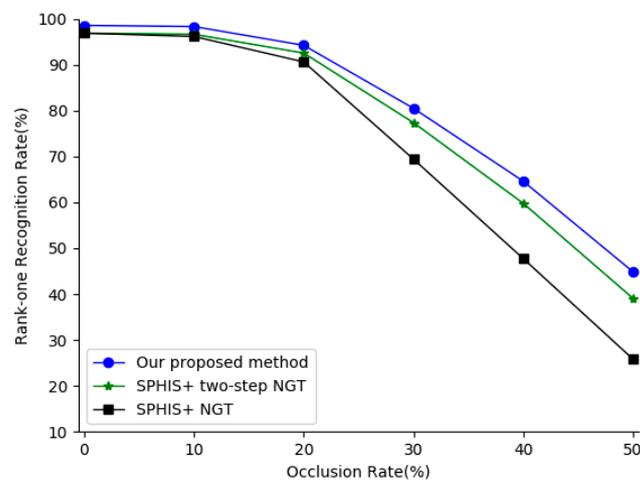


Figure 9. Recognition rates of our proposed method, SPHIS + two-step NGT, and SPHIS + NGT method for various random occlusion rates.

8. Conclusions

In this paper, we presented an automatic 3D ear biometric system using range images and applied it to ear recognition with partial occlusion. Within the system, a proposed spherical-window-based keypoint detection algorithm robustly detects 3D keypoints for the ear. The proposed 3D ear surface matching approach employs both the SPHIS feature-based local-feature matching method and the NGT-based matching method to obtain the keypoint correspondence. Furthermore, the proposed approach employs both the number of local keypoint correspondences and the holistic registration error. Extensive experiments on UND datasets with natural occlusion and various percentages of occlusion at random locations showed that the proposed method is accurate and robust against image occlusion.

Future more works will include investigating the use of the proposed 3D ear recognition system for recognition tasks involving general 3D objects with multiple randomly located occlusion blocks of various shapes.

Author Contributions: Q.Z. and Z.M. conceived and designed the experiments. Q.Z. performed the experiments. Q.Z. analyzed the data. Q.Z. wrote the paper. Q.Z. and Z.M. reviewed and edited the manuscript. All authors read and approved the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant Nos. 61375010), the National Natural Science Foundation of China (Grant Nos.61472031), the National Natural Science Foundation of China (Grant Nos.61300075).

Acknowledgments: This article was supported by the National Natural Science Foundation of China (Grant Nos. 61375010), the National Natural Science Foundation of China (Grant Nos.61472031), the National Natural Science Foundation of China (Grant Nos.61300075). The authors would like to thank the computer vision research laboratory at University of Notre Dame for providing their biometrics databases.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Emeršič, Ž.; Meden, B.; Peer, P.; Štruc, V. Evaluation and analysis of ear recognition models: Performance, complexity and resource requirements. In *Neural Computing and Applications*; Springer: Berlin, Germany, 2018; pp. 1–16.
2. Emeršič, Ž.; Štruc, V.; Peer, P. Ear recognition: More than a survey. *Neurocomputing* **2017**, *255*, 26–39. [[CrossRef](#)]
3. Ganapathi, I.I.; Prakash, S. 3D ear recognition using global and local features. *IET Biom.* **2018**, *7*, 232–241. [[CrossRef](#)]
4. Prakash, S. False mapped feature removal in spin images based 3D ear recognition. In Proceedings of the 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 11–12 February 2016; pp. 620–623.
5. Krotewicz, P. Novel ear-assisted 3D face recognition under expression variations. *Int. J. Biom.* **2016**, *8*, 65–81. [[CrossRef](#)]
6. Chen, H.; Bhanu, B.; Wang, R. Performance evaluation and prediction for 3D ear recognition. In Proceedings of the International Conference on Audio- and Video-Based Biometric, Rye Brook, NY, USA, 20–22 July 2005; pp. 748–757.
7. Maity, S.; Abdel-Mottaleb, M. 3D ear segmentation and classification through indexing. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 423–435. [[CrossRef](#)]
8. Zhang, Y.; Mu, Z.; Yuan, L.; Zeng, H.; Chen, L. 3D Ear Normalization and Recognition Based on Local Surface Variation. *Appl. Sci.* **2017**, *7*, 104. [[CrossRef](#)]
9. Albarelli, A.; Rodola, E.; Bergamasco, F.; Torsello, A. A non-cooperative game for 3d object recognition in cluttered scenes. In Proceedings of the 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, Hangzhou, China, 16–19 May 2011; pp. 252–259.
10. Rodolà, E.; Albarelli, A.; Bergamasco, F.; Torsello, A. A scale independent selection process for 3d object recognition in cluttered scenes. *Int. J. Comput. Vis.* **2013**, *102*, 129–145. [[CrossRef](#)]
11. Zhou, J.; Cadavid, S.; Abdel-Mottaleb, M. An efficient 3-D ear recognition system employing local and holistic features, *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 978–991. [[CrossRef](#)]
12. Ansari, S.; Gupta, P. Localization of ear using outer helix curve of the ear. In Proceedings of the International Conference on Computing: Theory and Applications, Kolkata, India, 5–7 March 2007; pp. 688–692.
13. Sarangi, P.P.; Panda, M.; Mishra, B.P.; Dehuri, S. An automated ear localization technique based on modified Hausdorff distance. In Proceedings of the International Conference on Computer Vision and Image Processing, Roorkee, India, 9–12 September 2017; pp. 229–240.
14. Zhang, Y.; Mu, Z. Ear Detection under Uncontrolled Conditions with Multiple Scale Faster Region-Based Convolutional Neural Networks. *Symmetry* **2017**, *9*, 53. [[CrossRef](#)]
15. Hurley, D.J.; Nixon, M.S.; Carter, J.N. Force field energy functionals for image feature extraction. *Image Vis. Comput.* **2002**, *20*, 311–317. [[CrossRef](#)]
16. Yuan, L.; Mu, Z.C. Ear detection based on skin-color and contour information. In Proceedings of the International Conference on Machine Learning and Cybernetics, Qingdao, China, 11–14 July 2010; pp. 2213–2217.
17. Arbab-Zavar, B.; Nixon, M.S. On shape-mediated enrolment in ear biometrics. In *Advances in Visual Computing*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 549–558.
18. Cummings, A.H.; Nixon, M.S.; Carter, J.N. A novel ray analogy for enrolment of ear biometrics. In Proceedings of the 2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), Washington, DC, USA, 27–29 September 2010; pp. 1–6.
19. Prakash, S.; Gupta, P. An efficient ear localization technique. *Image Vis. Comput.* **2012**, *30*, 38–50. [[CrossRef](#)]
20. Yan, P.; Bowyer, K.W. Biometric recognition using 3D ear shape. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *29*, 1297–1308. [[CrossRef](#)] [[PubMed](#)]

21. Deepak, R.; Nayak, A.V.; Manikantan, K. Ear detection using active contour model. In Proceedings of the International Conference on Emerging Trends in Engineering, Technology and Science, Pudukkottai, India, 24–26 February 2016; pp. 1–7.
22. Ganesh, M.R.; Krishna, R.; Manikantan, K.; Ramachandran, S. Entropy based Binary Particle Swarm Optimization and classification for ear detection. *Eng. Appl. Artif. Intell.* **2014**, *27*, 115–128. [[CrossRef](#)]
23. Halawani, A.; Li, H. Human ear localization: A template-based approach. In Proceedings of the International Workshop on Pattern Recognition (ICOPR 2015), Dubai, UAE, 4–5 May 2015; pp. 4–5.
24. Islam, S.M.; Bennamoun, M.; Davies, R. Fast and fully automatic ear detection using cascaded adaboost. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Copper Mountain, CO, USA, 7–9 January 2008; pp. 1–6.
25. Shih, H.C.; Ho, C.C.; Chang, H.T.; Wu, C.S. Ear detection based on arc-masking extraction and AdaBoost polling verification. In Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2009), Kyoto, Japan, 12–14 September 2009; pp. 669–672.
26. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Machine Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
27. Cintas, C.; Quinto-Sánchez, M.; Acuña, V.; Paschetta, C.; De Azevedo, S.; de Cerqueira, C.C.S.; Canizales-Quinteros, S. Automatic ear detection and feature extraction using geometric morphometrics and convolutional neural networks. *IET Biom.* **2016**, *6*, 211–223. [[CrossRef](#)]
28. Emersic, Z.; Gabriel, L.; Struc, V.; Peer, P. Convolutional Encoder-Decoder Networks for Pixel-wise Ear Detection and Segmentation. *IET Biom.* **2018**. [[CrossRef](#)]
29. Chen, H.; Bhanu, B. Human ear recognition in 3D. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 718–737. [[CrossRef](#)] [[PubMed](#)]
30. Cadavid, S.; Abdel-Mottaleb, M. 3-D ear modeling and recognition from video sequences using shape from shading. *IEEE Trans. Inform. Forensics Secur.* **2018**, *3*, 709–718. [[CrossRef](#)]
31. Islam, S.M.; Davies, R.; Bennamoun, M.; Mian, A.S. Efficient detection and recognition of 3D ears. *Int. J. Comput. Vis.* **2011**, *95*, 52–73. [[CrossRef](#)]
32. Zhang, L.; Ding, Z.; Li, H.; Shen, Y. 3D ear identification based on sparse representation. *PLoS ONE* **2014**, *9*, e95506. [[CrossRef](#)] [[PubMed](#)]
33. Zhou, J.; Cadavid, S.; Abdel-Mottaleb, M. A computationally efficient approach to 3d ear recognition employing local and holistic features. In Proceedings of the 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Colorado Springs, CO, USA, 20–25 June 2011; pp. 98–105.
34. Liu, Y.; Zhang, B.; Lu, G.; Zhang, D. Online 3D Ear Recognition by Combining Global and Local Features. *PLoS ONE*, **2016**, *11*, e0166204. [[CrossRef](#)] [[PubMed](#)]
35. Ganapathi, I.I.; Ali, S.S.; Prakash, S. Geometric statistics-based descriptor for 3D ear recognition. In *The Visual Computer*; Springer: Berlin, Germany, 2018; pp. 1–13.
36. Tian, L.; Mu, Z. Ear recognition based on deep convolutional network. In Proceedings of the International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Datong, China, 15–17 October 2016; pp. 437–441.
37. Chowdhury, M.; Islam, R.; Gao, J. Robust ear biometric recognition using neural network. In Proceedings of the 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), Siem Reap, Cambodia, 18–20 June 2017; pp. 1855–1859.
38. Omara, I.; Wu, X.; Zhang, H.; Du, Y.; Zuo, W. Learning pairwise SVM on deep features for ear recognition. In Proceedings of the IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 24–26 May 2017; pp. 341–346.
39. Emeršič, Ž.; Štepec, D.; Štruc, V.; Peer, P. Training convolutional neural networks with limited training data for ear recognition in the wild. In Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG), Washington, DC, USA, 30 May–3 June 2017; pp. 987–994.
40. Almisreb, A.A.; Jamil, N.; Din, N.M. Utilizing AlexNet Deep Transfer Learning for Ear Recognition. In Proceedings of the Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), Kota Kinabalu, Malaysia, 26–28 March 2018; pp. 1–5.
41. Zhang, Y.; Mu, Z.; Yuan, L.; Yu, C. Ear verification under uncontrolled conditions with convolutional neural networks. *IET Biom.* **2018**, *7*, 185–198. [[CrossRef](#)]

42. Hansley, E.E.; Segundo, M.P.; Sarkar, S. Employing fusion of learned and handcrafted features for unconstrained ear recognition. *IET Biom.* **2018**, *7*, 215–223. [[CrossRef](#)]
43. Ying, T.; Shining, W.; Wanxiang, L. Human ear recognition based on deep convolutional neural network. In Proceedings of the Chinese Control and Decision Conference (CCDC), Shenyang, China, 9–11 June 2018; pp. 1830–1835.
44. Yaman, D.; Eyiokur, F.I.; Sezgin, N.; Ekenel, H.K. Age and Gender Classification from Ear Images. In Proceedings of the International Workshop on Biometrics and Forensics (IWBF), Sassari, Italy, 7–8 June 2018; pp. 1–7.
45. Fan, T.Y.; Mu, Z.C.; Yang, R.Y. Multi-modality recognition of human face and ear based on deep learning. In Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), Ningbo, China, 9–12 July 2017; pp. 38–42.
46. Garcia-Garcia, A.; Gomez-Donoso, F.; Garcia-Rodriguez, J.; Orts-Escolano, S.; Cazorla, M.; Azorin-Lopez, J. Pointnet: A 3D convolutional neural network for real-time object class recognition. In Proceedings of the International Joint Conference on Neural Networks, Vancouver, BC, Canada, 24–29 July 2016; pp. 1578–1584.
47. Qi, C.R.; Su, H.; Nießner, M.; Dai, A.; Yan, M.; Guibas, L.J. Volumetric and multi-view cnns for object classification on 3D data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas Valley, NV, USA, 26 June–1 July 2016; pp. 5648–5656.
48. Xu, X.; Corrigan, D.; Dehghani, A.; Caulfield, S.; Moloney, D. 3D Object Recognition Based on Volumetric Representation Using Convolutional Neural Networks. In Proceedings of the International Conference on Articulated Motion and Deformable Objects, Palma de Mallorca, Spain, 13–15 July 2016; pp. 147–156.
49. Klovov, R.; Lempitsky, V. Escape from cells: Deep kdnetworks for the recognition of 3D point cloud models. In Proceedings of the IEEE International Conference on Computer Vision, Copenhagen, Denmark, 7 September–11 August 2017; pp. 863–872.
50. Feng, Y.; Zhang, Z.; Zhao, X.; Ji, R.; Gao, Y. GVCNN: Group-View Convolutional Neural Networks for 3D Shape Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 264–272.
51. Chen, L.; Mu, Z. Partial data ear recognition from one sample per person. *IEEE Trans. Hum.-Mach. Syst.* **2016**, *46*, 799–809. [[CrossRef](#)]
52. Koenderink, J.J.; Van Doorn, A.J. Surface shape and curvature scales. *Image Vis. Comput.* **1992**, *10*, 557–564. [[CrossRef](#)]
53. Dorai, C.; Jain, A.K. COSMOS-A representation scheme for 3D free-form objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 1115–1130. [[CrossRef](#)]
54. Jolliffe, I. *Principal Component Analysis*; Wiley: Hoboken, NJ, USA, 2005.
55. Bulò, S.R.; Bomze, I.M. Infection and immunization: A new class of evolutionary game dynamics. *Games Econ. Behav.* **2011**, *71*, 193–211. [[CrossRef](#)]
56. Turk, G.; Levoy, M. Zippered polygon meshes from range images. In Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, Orlando, FL, USA, 24–29 July 1994; pp. 311–318.
57. Albarelli, A.; Rodola, E.; Torsello, A. A game-theoretic approach to fine surface registration without initial motion estimation. In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010; pp. 430–437.
58. Rusinkiewicz, S.; Levoy, M. Efficient variants of the ICP algorithm. In Proceedings of the Third International Conference on 3-D Digital Imaging and Modeling, Quebec City, QC, Canada, 28 May–1 June 2001; pp. 145–152.
59. Cappelli, R.; Maio, D.; Maltoni, D. Combining fingerprint classifiers. In Proceedings of the International Workshop on Multiple Classifier Systems, Cagliari, Italy, 21–23 June 2000; pp. 351–361.

