

Article

Biometric Security: A Novel Ear Recognition Approach Using a 3D Morphable Ear Model

Md Mursalin , Mohiuddin Ahmed *  and Paul Haskell-Dowland 

School of Science, Edith Cowan University, Perth 6027, Australia

* Correspondence: mohiuddin.ahmed@ecu.edu.au

Abstract: Biometrics is a critical component of cybersecurity that identifies persons by verifying their behavioral and physical traits. In biometric-based authentication, each individual can be correctly recognized based on their intrinsic behavioral or physical features, such as face, fingerprint, iris, and ears. This work proposes a novel approach for human identification using 3D ear images. Usually, in conventional methods, the probe image is registered with each gallery image using computational heavy registration algorithms, making it practically infeasible due to the time-consuming recognition process. Therefore, this work proposes a recognition pipeline that reduces the one-to-one registration between probe and gallery. First, a deep learning-based algorithm is used for ear detection in 3D side face images. Second, a statistical ear model known as a 3D morphable ear model (3DMEM), was constructed to use as a feature extractor from the detected ear images. Finally, a novel recognition algorithm named you morph once (YMO) is proposed for human recognition that reduces the computational time by eliminating one-to-one registration between probe and gallery, which only calculates the distance between the parameters stored in the gallery and the probe. The experimental results show the significance of the proposed method for a real-time application.

Keywords: ear biometrics; detection; 3D morphable model; recognition



Citation: Mursalin, M.; Ahmed, M.; Haskell-Dowland, P. Biometric Security: A Novel Ear Recognition Approach Using a 3D Morphable Ear Model. *Sensors* **2022**, *22*, 8988. <https://doi.org/10.3390/s22228988>

Academic Editor: Loris Nanni

Received: 9 October 2022

Accepted: 17 November 2022

Published: 20 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Biometrics is a critical component of cybersecurity that identifies persons by verifying their behavioral and physical traits. It is the most precise and powerful physical security solution for identity verification presently in use. In biometric-based authentication, each individual can be correctly recognized based on their intrinsic behavioral or physical features. Biometrics systems based on physiological attributes such as the face [1], fingerprint [2], iris [3], palm prints [4], and ear [5] have been described by researchers [6]. Examples of these physiological biometrics qualities are shown in Figure 1. If someone tries to get into the biometric security system, it scans them, analyses their traits, and compares them to previously recorded information. The individual is given access to the facility or equipment when there is a match found.

Biometrics can be categorized into touch-based and touchless. The fingerprint is one of the most popular touch-based biometrics used presently. However, with the advancement of technology, hackers can mimic fingerprints and access important information. Moreover, the weather conditions such as rain, snow, or humidity can cause problems with the fingerprint-based security system. Furthermore, the quality of the hardware degrades over time. Apart from touch-based, facial recognition is considered the prominent touchless biometrics. Although it shows advantages over fingerprints, it faces numerous challenges, including illumination, and poses variations, ages, etc. A sample challenge in the facial recognition system is illustrated in Figure 2, where the images (probe) of the same person show lots of deviation from the original stored image (gallery). Therefore, the face-based approach has to be robust against those situations which infer the need for additional biometric traits. The human ear is considered a crucial biometric, showing

highly distinguishing features. Researchers found that even identical twins have different ear shapes [7]. The advantage of ear image analysis over other biometric traits, such as faces, iris, fingerprints, and palm prints can be attributed to its ease of capture, invariance to expressions, and stability over time [8]. Due to these benefits, ear images can be used for various purposes such as biometric identification, clinical asymmetry research, genetic connection investigation, and gender recognition [9–12].

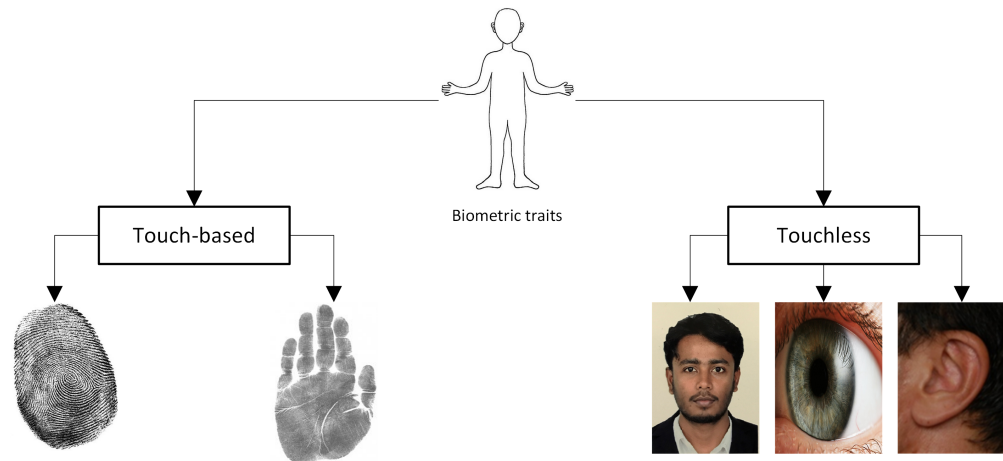


Figure 1. Various biometric traits for human recognition.

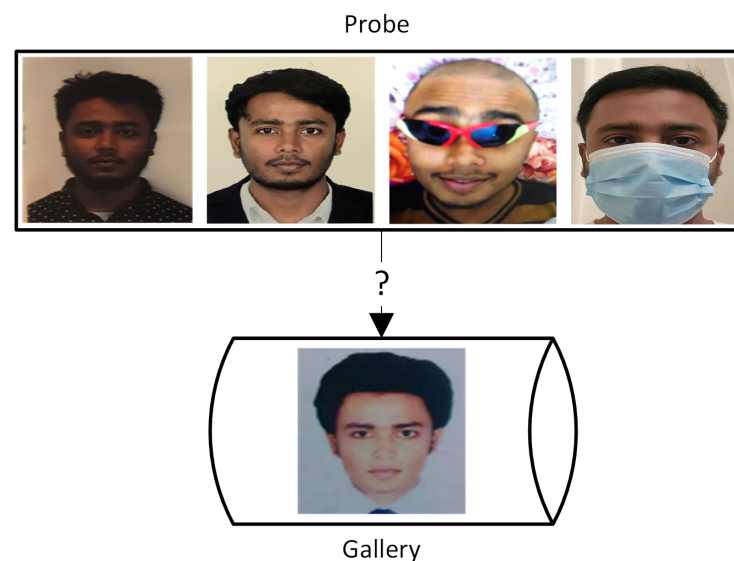


Figure 2. Face recognition challenges. All of these images (same person) are taken within 2–3 years.

Generally, touch-based biometrics such as fingerprints and palm prints should be avoided for public safety due to the rising concern about the COVID-19 pandemic. As a result, a touchless biometric system for person recognition in real-world applications such as office attendance, access management, banking, and surveillance is in high demand. Faces are likewise non-intrusive biometrics but have a barrier because they are often hidden behind masks. Due to its non-intrusive nature, the ear is a valuable biometric in this situation where a facial mask covers the face, but the ear region is visible.

Although 2D ear image analysis is more widespread because of its ease of computation, it has certain limitations, such as sensitivity to illumination and poses variations. Moreover, the shape information is limited for a small object, such as an ear. On the contrary, a 3D image provides much more shape information, even for a small object. Therefore, 3D ear image analysis can be a promising prospect for human recognition.

In an ear-based recognition system, a gallery is referred to as a set of images indicating all individuals known to the system. This work provides one image per person to create the gallery. The images of the same individual not in the gallery are called probe images. In this work, we evaluated two recognition tasks. The first task is identification, where the system notices which person from the gallery is revealed on the probe image. The second task is verification, where a person declares to be a certain gallery image. The system determines if the probe and the gallery image show the same person.

Usually, in the literature, each probe is used to register using the registration algorithm for all the gallery images. Therefore, it is very time-consuming and computationally expensive. This work develops an algorithm named YMO that needs only one registration to deform the 3DMEM towards a given query image and calculate the shape parameters. After computing the parameters, this approach only calculates the distance between the parameters stored in the gallery and the probe without requiring further registration steps. The recognition result is verified based on a threshold. The contribution of this work can be summarised as follows,

- A novel recognition algorithm named YMO is developed to match the probe and gallery, reducing computational time.
- The performance of different distance metrics for ear recognition is demonstrated.
- A comparative study shows the comparable performance of our 3D ear recognition method.

2. Literature Review

The existing 3D ear recognition method mostly used either local features or global features for representing the feature space. These features were extracted after pairwise registering the ears. As a result, the recognition process becomes computationally heavy when a given probe image needs to register with all the gallery images. Before 2005, existing ear recognition methods did not show any performance prediction, either theoretically or experimentally. The existing 3D ear recognition method can be categorized based on feature types such as local, global, and fusion of local and global features. This section briefly explains the different existing methods that use these features.

2.1. Local Feature

Chen et al. [13] first proposed a new representation of a 3D ear based on an integrated local feature to recognize human ears. They defined the features using local minimum and maximum shape indexes from principal curvatures. An initial correspondence of the local surface patches was established by comparing local surface patches between an offline model and test images. The performance of their method was evaluated on 52 subjects using the cumulative match characteristic (CMC) curve. Later, they [14,15] introduced a method for ear recognition by combining feature embedding and support vector machine (SVM). The local surface patch representation established the correspondence between the model and test images. To reduce the high dimensionality of the feature vector, they applied a feature embedding algorithm. The similarity between a model and test pair was calculated by searching the nearest neighbors from the low dimensional embedded features. The similarities for all model and test pairs were ranked using the SVM algorithm that generates a shortlist of candidate models for verification.

To encode 2D and 3D local features for ear recognition, Chen et al. [16] introduced the texture and depth scale-invariant feature 100 transform (TDSIFT). The TDSIFT showed its superiority over the conventional scale-invariant feature transform (SIFT) descriptor by fusing 2D and 3D local information. They applied a key-point detection method on 2D images and then projected the corresponding key point on the co-registered 3D images to form the TDSIFT descriptor.

By identifying key points that utilize the curvilinear structure in 2D ear pictures projected to the co-registered 3D ear images, Ganapathi et al. [17] developed a local feature description-based technique for human recognition. A feature descriptor vector was calcu-

lated from the neighborhood around each mapped key point in 3D. The correspondence between each pair of probe and gallery images was established by using the Iterative Closest Point (ICP) algorithm [18]. The registration score was used to determine the matching score. Later they [19] proposed a feature extraction method using geometric statistics that eliminates the dependency on 2D co-registered images. However, to enhance the feature extraction, the authors [20] extracted additional descriptors by utilizing neural network-based auto-encoders and local statistics of the depth images. In addition, the authors [18] proposed a multi-modal approach that used 2D images to identify key points and co-registered 3D images to extract features from the key points. They analyzed six key-point detectors with the ICP algorithm and reported the performance.

A 3D local feature-based (L3DF) method that merged the ear and face for human recognition was proposed by Islam et al. [21]. They detected key points based on asymmetrical variations in depth from both faces and ears. An altered ICP was applied to align the probe and gallery dataset. They utilized the weighted sum during the fusion to weigh more on face features over ear data. The recognition performance for their work was mostly dependent on the face. Therefore, the authors [22] proposed a method for 3D ear recognition that did not require facial features. They applied an AdaBoost algorithm to detect ears from 2D profile face images that were projected to the co-registered 3D images to crop the 3D ear. A speeded-up robust feature (SURF) feature-based method was proposed by Prakash et al. [23]. The SURF features were extracted from the co-registered 3D images.

To extract local features directly on 3D ear point clouds, Sun et al. [24] proposed a method using a Gaussian-weighted average of the mean curvature of each point. They presented an optimal selection of the salient key points utilizing the Poisson disk sampling. Subsequently, the authors created a local feature descriptor of each salient key point by fitting a surface to the neighborhood of each salient key point using the quadratic principal manifold method. Later, the authors [25] investigated the local shape feature combined into the joint α -entropy of the minimum spanning tree (MST). They used their previous technique to detect salient points from the 3D ear image and fit the neighborhood of each salient point to a single-value quadric surface. Each salient point's local shape feature vector was specified as the sampling depth set on the parametric node of the quadric surface. They constructed the MST on the matched key points for every pair of gallery ears and probe ears. Eventually, the authors reduced the total edge weight of MST to determine the similar pair if its joint α -entropy value is small.

A method using shape information to arrange the 3D ear data in a hierarchical categorization was presented by Maity et al. [26], where the ear was segmented by applying an active contour algorithm with a tree-structured graph. These 3D ears were partitioned into different categories based on geometric shapes, including round, oval, rectangular, etc. They used indexing methods with balanced split (KD tree) and unbalanced split (pyramid tree) data structures to distinctly categorize the database. Zeng et al. [27] presented a similar shape-based method. They created the 3D Center-Symmetric Local Binary Patterns (CS-LBP) features and utilized a coarse-to-fine approach for 3D salient point matching. The matching scores were computed using the average Earth mover's distance (EMD) distances for 3D ear recognition. Later the authors [28] improved the method by using a modified iterative closest point (MICP). The authors computed three descriptors: LBP descriptor, 3D LBP descriptor, and 3D CS-LBP descriptor for feature extraction and matching. For leading additional local feature information into global registration, Zhang et al. [29] used a one-step ICP local surface variation (LSV) algorithm for a 3D ear matching scheme. They applied data normalization to eliminate the background noise.

2.2. Global Feature

By generating 3D data from video frames, Cadavid et al. [30] suggested a technique for ear recognition. For 3D ear identification systems, they used two algorithms called shape from shading (SFS) and structure from motion (SFM). The ear region was segmented from each frame in a video sequence using interpolation of ridges and ravines. They

reconstructed 3D shapes by tracking key points across video frames and using the factorization method. This reconstructed 3D ear model was matched using ICP for recognition. Later they [31] increased the number of subjects and applied the SFS method for recognition. However, they did not perform any robustness against poses or occlusions. A similar video-based approach was proposed by Mahoor et al. [32] for multi-modal face and ear recognition. They reconstruct 3D images from a series of video frames using the SFS method. The 3D models were registered using the ICP algorithm. The Active Shape Model was used to derive a collection of facial landmarks from frontal facial photos for 2D face recognition. Then, at the positions of face landmarks, the response of facial images to a sequence of Gabor filters is determined. The best match was determined by comparing the Gabor features of a probe face image to those of the reference models. The ear recognition and face recognition modality match scores are combined to improve the system's overall recognition rate.

A 3D ear recognition technique was presented by Passalis et al. [33] utilizing a general annotated ear model (AEM). This AEM was registered and placed into each ear to generate a biometric signature with 3D data storage. After the registration process, depth and normal images were generated and concatenated to form the biometric signature coefficients. The L^1 metric was used to compare the coefficients between two images. For further improvement, the authors [34] proposed a unified method that fuses 3D facial and ear data. They used an annotated deformable model fitting to the data and calculated wavelet coefficients from the geometry image for the biometric signature.

A theoretical approach for calculating the similarity of equivalent feature sets was presented by Tre et al. [35]. Their research aimed to demonstrate how computational intelligence could improve ear recognition. They proposed bipolar data modeling and aggregation methods to represent the data to improve the performance against noises. The similarity between the two ear images was measured by calculating the Minkowski distance and managing the hesitation caused by poor image quality [36]. They also proposed a hierarchically structured comparison method for features.

3. Proposed Method

This work proposed a complete pipeline for ear recognition from 3D profile images. The profile is considered a left-side 3D image in this work. The first step is to detect and extract the ear from the profile face image. This work used a deep learning-based method for ear detection that can detect ears from the 3D point cloud representation of the profile face. The next step is to create a 3DMEM from the extracted 3D ear images. A combination of rigid ICP and a variant of a non-rigid ICP algorithm are used to register two ear images in a way so that each point of one ear should correspond to the other ear. This representation is known as dense correspondence. After achieving dense correspondence, we can use linear representation to apply statistical analysis, which facilitates the creation of a statistical shape model known as the 3D morphable ear model (3DMEM). This 3DMEM can be used as a generic representation of the ear shape and can be used for parameterization for a new instance. Next, a novel matching technique is proposed to extract the shape parameters. This work demonstrates various distance metrics to find the best solution. The recognition is considered successful when the distance value satisfies the threshold condition. The proposed ear recognition method is illustrated in Figure 3.

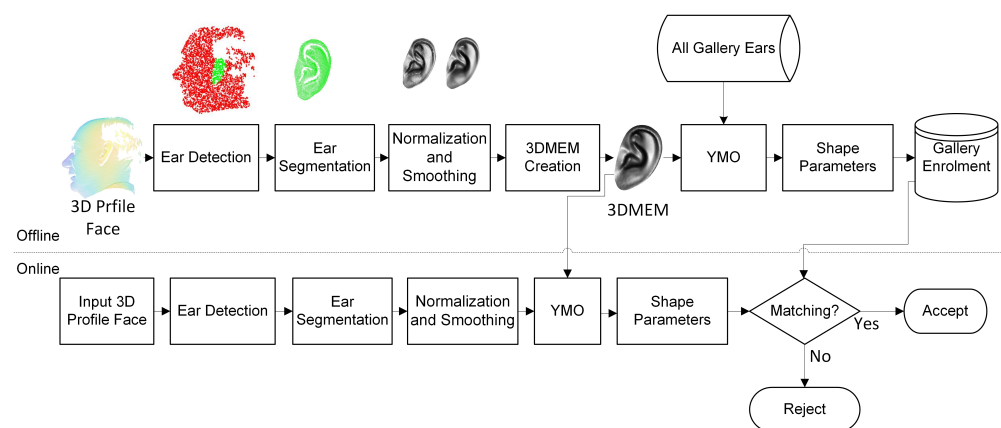


Figure 3. Block diagram of the proposed ear recognition method.

3.1. Dataset Preparation

The proposed technique was validated on the University of Notre Dame (UND) J2 database. This database is considered one of the most extensively accessible ear databases, with 1800 samples from 415 subjects. The images in the database are affected by pose changes, scaling, and occlusions due to earrings and hair. The recognition system requires at least one gallery image per person. In this work, we only considered subjects with two or more samples, so the gallery should contain one image per subject. Among 415 subjects of the UND J2 dataset, we found 404 subjects that satisfied our condition. For ear recognition, the first task is to create a gallery and probe dataset from the chosen 404 subjects with 1780 samples. A gallery dataset consists of 404 arbitrarily chosen images from each subject, and the probe comprises the remaining 1376 images.

3.2. Ear Detection

The ear detection from 3D profile faces is performed by adopting our previously developed fully automated deep learning-based algorithm named EarNet [37]. The EarNet is a modified version of the PointNet++ [38] architecture. EarNet is lighter than PointNet++, making the computation significantly faster. Additionally, a data augmentation block was included to rotate the full 3D ear point cloud. This rotation was performed with respect to the x and y axes. This augmentation leverages the understanding of a given profile face object and enhances the performance of ear detection in 3D point clouds. The EarNet is trained from scratch utilizing 20,000 synthetic data (training and testing split was 80% and 20%). We empirically selected the hyperparameters. The experimental observation showed the optimal batch size was 16. We assigned the number of data points of each scan as 4096. The optimizer was set as Adam [39] with a momentum of 0.9. We choose the initial learning rate as 10^{-3} . To improve the robustness, we used transfer learning to the network utilizing 150 real 3D scans (arbitrarily selected from the UND J2 dataset). We applied rotation augmentation during transfer learning. All experiments were conducted in the Lambda Blade machine with GPU 8 × 1080 Ti GeForce GTX 1080 Ti. The EarNet shows 100% detection accuracy in the UND J2 dataset. More details can be found in [37,40].

3.3. Morphable Model Generation

After extracting the ear from a 3D profile face image, each ear image is parameterized so we can apply statistical operations [41]. This parameterization is achieved by morphing a template to every ear in the database. This is an iterative process where the template is deformed using a non-rigid ICP algorithm [42]. To select a template, we first randomly chose one sample ear from the ear database and registered it with all the remaining ears. After registering, we calculate the mean ear. This mean ear is again used as a new template, and we continue the same procedure again by registering all the ear image with the new template. So the output of the registration step is the set of reparameterized ear images

where each point of the ear semantically corresponds with the other ears. Therefore, we can apply linear operations to these registered images. Now it is possible to create a 3DMEM using these densely corresponded ears. We reduce the dimensionality by using principal component analysis to construct the 3DMEM. By varying the shape parameters of the 3DMEM, we can generate novel ear-shaped instances. Let, the matrix of all densely corresponded ears $\Gamma = [v_1, v_2, \dots, v_N]$, where $v = [x_1, \dots, x_n, y_1, \dots, y_n, z_1, \dots, z_n]^T$ and number of vertices $n = 1, \dots, N$. Now, the statistical shape model (M) can be expressed as,

$$M_i = \mu + \sum_{i=1}^d \alpha_i U_i = \mu + \alpha U \quad (1)$$

We can calculate the mean shape (μ) from the densely corresponded ears using the following equation,

$$\mu = \sum_{i=1}^N e_i \quad (2)$$

The row-normalized matrix R is computed by subtracting the μ from each 3D ear image.

$$R = \Gamma - \mu \quad (3)$$

The eigenvalue is computed by applying a singular value decomposition method named *svd*. Here, U stands for principal components, S represents the diagonal matrix of eigenvalues, and V implies the corresponding loading.

$$[U, S, V] = \text{svd}(R) \quad (4)$$

where α represents the shape parameters, which are applied to alter the shape, and d is the number of principal components. The shape parameter α is computed utilizing the following equation,

$$\alpha = U^T (m - \mu) \quad (5)$$

3.4. You Morph Once Algorithm

The YMO algorithm starts with a rigid registration process followed by shape deformation. In this work, we adopted a fitting algorithm proposed by [43] for face deformation. However, we updated the method with different distance calculations for ear model deformation. The purpose is to deform the mean ear shape from the 3DMEM to minimize the distance between the query ear QE and the deformed mean ear EM . The query ear after vectorization is parametrized by the statistical model as $qm_i = U\alpha_i + \mu Y$, where the vector α_i consists of parameters. These parameters can vary the model's shape in the i th iteration. The α_i is set to zero, and the mean ear of the 3DEM characterizes the deformable model EM_i in the initialization step. Each iteration starts with a registration step where the input ear QE is registered to the model EM_i . In this step, an approximate correspondence between the EM_i and the QE is computed with a rigid transformation. The correspondence is determined by analyzing the Nearest Neighbor (NN) of each point of EM_i in QE following a k-d tree data structure. Suppose d represents the Spearman distance (see section Distance metrics) between the corresponding query ear and the model. The outliers are defined as points on QE' whose NN distance with EM_i is more significant than a threshold th where $th = \bar{d} + 2\sigma d$ and remove them from registration. This measure confirms that outliers do not influence the registration process. Next, the QE is translated to the μ of the model and is rotated to align with EM_i . Here, QEr denotes the corresponding and registered query ear. In the following step, the model EM_i is deformed to fit the registered query ear QEr as

$$\alpha_i = \min_{\alpha_i} \|Me - qe_r\|_2 + \lambda \|\alpha_i - \alpha_{i-1}\|_2 \quad (6)$$

where $Me = U * \alpha_i + \mu \gamma$, is calculated from the morphable ear model. The iterative process is completed when the residual error (η) between qm_i and qe_r is less than equal to 10^{-5} . A stiffness weight λ was used to regularize the shape. In this experiment the λ was set to 0.6.

After deforming the 3DMEM to the probe, the last step is to calculate the distance between the deformed 3DMEM with all of the gallery shape parameters. In this work, the Spearman distance is used to find the distance. The minimum distance is considered as the matching pair between the probe and the gallery. The pseudocode is shown in Algorithm 1.

Algorithm 1 YMO Algorithm

Require: 3DMEM, QE, GAL
Ensure: QE is upward facing
 $EM_0 = \mu$
 $i = 1$
while $e_i > \eta$ **do**
 $QE' = dist_{nm}(QE, EM_i) < th$
 $QEr = rigidICP(QE', EM_i)$
 $Me = U * \alpha_i + \mu\gamma$
 $\alpha_i = min_{\alpha_i} ||Me - qe_r||_2 + \lambda ||\alpha_i - \alpha_{i-1}||_2$
 $qm_i = U\alpha_i + \mu Y$
 $e_i = dist_{sp}(m^i - q_r)$
 $i++$
end while
 Calculate the SP distance between DEM and GAL
 Find the min(SP)
if min(SP) < thVal **then**
 Match found in the gallery.
else
 Not matched!
end if

3.5. Gallery Enrollment

The first task for any recognition is to create a gallery database of each individual to match the probe image. This work uses a novel approach to create the gallery database where a fitting algorithm is utilized to deform the 3DMEM as close to a given individual ear image. After the deformation of the 3DMEM, the parameters are calculated and stored in the gallery database. These parameters are used to transform the 3DMEM toward the given input image.

3.6. Evaluation Metrics

3.6.1. Distance Metrics

A distance metric is a function that shows how far apart two observations are from one another. This study employs a number of different distance measures, including the Euclidean and normalized Euclidean distances, as well as the Mahalanobis, Minkowski, city block, Chebyshev, cosine, correlation, Spearman, and Jaccard distances. Given a data matrix X that is represented by $m(1 \times n)$ row vectors x_1, x_2, \dots, x_m . As well as a $m \times n$ data matrix Y , which is represented as $m(1 \times n)$ row vectors y_1, y_2, \dots, y_m . The following definitions describe the distances between the vectors x_s and y_t ,

Euclidean distance (E):

$$E = \sqrt{(x_s - y_t)(x_s - y_t)'} \quad (7)$$

Standardized Euclidean distance (SE):

Here, V is the $n \times n$ diagonal matrix where $(S(j))^2$ is the j th diagonal element, where for each dimension, S represents a vector of scaling factors.

$$SE = (x_s - y_t)V^{-1}(x_s - y_t)' \quad (8)$$

Mahalanobis distance (MH):

$$MH = (x_s - y_t)C^{-1}(x_s - y_t)' \quad (9)$$

where C represents the covariance matrix.

Minkowski distance (MN):

$$MN = \sqrt[p]{\sum_{j=1}^n |x_{sj} - y_{tj}|^p} \quad (10)$$

City block distance (CT):

$$CT = \sum_{j=1}^n |x_{sj} - y_{tj}| \quad (11)$$

Chebychev distance (CH):

$$CH = \max_j |x_{sj} - y_{tj}| \quad (12)$$

Cosine distance (CS):

$$CS = 1 - \frac{x_s y_t'}{\sqrt{(x_s x_s') (y_t y_t')}} \quad (13)$$

Correlation distance (CR):

$$CR = 1 - \frac{(x_s - \bar{x}_s)(y_t - \bar{y}_t)'}{\sqrt{(x_s - \bar{x}_s)(x_s - \bar{x}_s)' (y_t - \bar{y}_t)(y_t - \bar{y}_t)'}} \quad (14)$$

Spearman distance (SP):

$$SP = 1 - \frac{(\zeta_s - \bar{\zeta}_s)(\zeta_t - \bar{\zeta}_t)'}{\sqrt{(\zeta_s - \bar{\zeta}_s)(\zeta_s - \bar{\zeta}_s)' (\zeta_t - \bar{\zeta}_t)(\zeta_t - \bar{\zeta}_t)'}} \quad (15)$$

where ζ_s and ζ_t represent the coordinate-wise rank vectors of x_s and x_t .

3.6.2. Identification Metrics

For ear recognition, the shape is represented by a set of coefficients, aka shape parameters $\alpha = (\alpha_1, \dots, \alpha_n)^T$. Several distance metrics are used to evaluate the comparison between the gallery and probe ears e_1 and e_2 .

The performance of the proposed approach is evaluated using the identification rate, and verification accuracy, where verification accuracy is defined as follows:

$$\text{Verification Accuracy} = (100 - \frac{FAR + FRR}{2}) \quad (16)$$

where the false acceptance rate (FAR) estimates the risk that an unauthorized person would be mistakenly accepted by the system, and the false rejection rate (FRR) calculates the likelihood that a person with authorization will be mistakenly rejected by the system. A threshold value is used to determine the best values for FRR and FAR. Any change in the threshold's value will directly affect FAR and FRR. To determine the verification accuracy, FRR and FAR are combined to their optimal value. When the verification accuracy is the highest, we have determined that the FRR and FAR combination is the best option for the suggested approach.

4. Results and Discussion

The experiment on the UND J2 dataset shows 98.51% rank-1 accuracy. The cumulative match characteristic curve (CMC) for the top 20 ranks on the UND J2 dataset is illustrated in Figure 4.

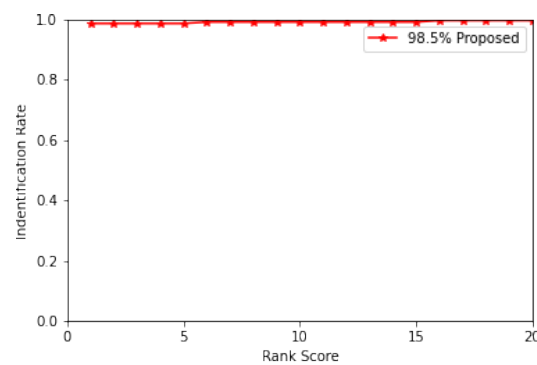


Figure 4. Identification rate on the UND J2 dataset.

The top-performing distance metric was SP distance. A comparison among different distance metrics is shown in Table 1. The reported results are based on one sample per person, where the probes were selected randomly during this experiment. The SP distance metrics utilize Spearman's correlation (SC) to rank the observations. Generally, the SC is suitable when the continuous data do not follow a line and have a monotonic relationship or ordinal data. In the case of a monotonic relationship, when one variable increases, the other variable shows either an increase or a decrease. Furthermore, this relationship does not have to be necessarily in a straight line. This aspect of Spearman's correlation allows fitting curvilinear relationships. As our extracted feature data points are continuous with a monotonic relationship and ranked using Spearman's correlation, SP distance metrics exhibit better results than other metrics.

We empirically found the threshold value 0.62, which provides the best results to verify a person as authentic or an imposter. The verification accuracy was 97% (according to Equation (16)). We have also measured the verification performance in terms of the receiver operating characteristic curve (ROC), which is shown in Figure 5.

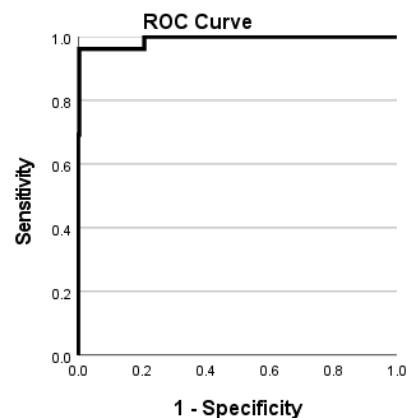


Figure 5. Performance evaluation of the proposed method on the UND J2 dataset using the ROC curve.

The proposed method performed well in the presence of natural hair near the ear region. Furthermore, we found that the proposed technique is able to recognize a person in the presence of earrings. Examples of ear recognition in the presence of earrings are illustrated in Figure 6. Additionally, we also demonstrate the performance by adding synthetic noise in the ear images. We found that the recognition performance reduces when adding more than 40% of noise.

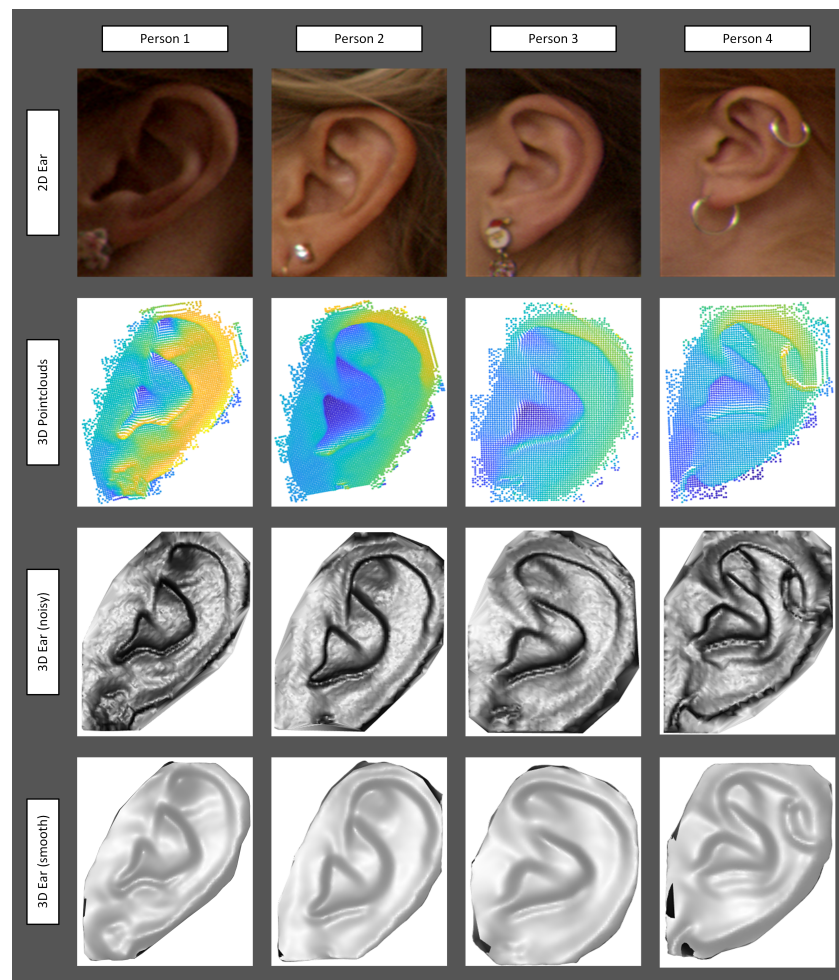


Figure 6. Examples of occlusions by earrings.

Table 1. Accuracy comparison among different distance metrics.

Distance Metrics	Correct/Wrong	Rank-1 Accuracy (%)
CH	273/131	67.57
MN	347/57	85.89
SE	347/57	85.89
E	347/57	85.89
CR	351/53	86.88
MH	351/53	86.88
CS	352/52	87.13
CT	381/23	94.31
SP	398/6	98.51

Compared to the other recognition systems shown in Table 2, the ear recognition system from this article exhibits greater identification accuracy. All the other techniques used one-to-one matching using registration-based algorithms. As a result, the computational time increase for a large gallery dataset. On the other hand, our proposed YMO algorithm deforms the 3DMEM to a given probe for extracting the shape parameters. After obtaining the parameters, it does not require additional one-to-one registration between the gallery and the proof; rather, it only calculates the distance from the gallery of shape parameters. Therefore, the computation time is not increased even for larger gallery datasets.

Table 2. The performance comparison in terms of rank-1 recognition accuracy of the proposed technique with existing state-of-the-art techniques in the literature.

Authors	Recognition Approach	Identification Rate (%)
Islam et al. [22]	L3DF and ICP	93.50
Prakash et al. [23]	SURF and GPA(ICP)	98.30
Yan et al. [44]	ICP	97.80
Sun et al. [25]	Key-point matching	95.1
Chen et al. [15]	LSP and ICP	96.36
This work	3DMEM and YMO	98.51

As the algorithms were applied in various hardware settings, the time cost in the table is simply for reference. Nevertheless, it is possible to gauge these algorithms' effectiveness using their computational complexity. Each ICP registration has a computational cost of $O(Iter * N_p \log_2(N_g))$, where $Iter$ is the number of iterations, N_p is the number of probe data points, and N_g is the number of gallery ear data points. The YMO algorithm converges to the minimum distance with fewer iterations when using the proposed recognition approach, which lowers the ear data to one-third the size of coarsely extracted ear area data. Therefore this ear recognition method is more efficient due to its reduced computational complexity. The average computation time to perform the recognition through a pairwise matching between probe and gallery templates for [26] was around 0.0039 s, while our method requires 0.0035 s. At the same time, [22] took 2.28 s for each probe and gallery pair, which is significantly large when the gallery data are increased. The computation time comparison is shown in Table 3.

Table 3. The comparison of computation time for the recognition phase of the proposed technique with existing state-of-the-art methods.

Authors	Mean Computation Time/Probe (s)
Islam et al. [22]	2.28
Jindan et al. [45]	0.019
Maity et al. [26]	0.0039
This work	0.0035 ± 0.0002

5. Conclusions

In this work, a novel approach based on a 3D morphable model is proposed for human recognition using ear images. For identification, all gallery images were analyzed by the fitting algorithm named YMO, and the shape coefficients were stored. For a given probe image, the YMO algorithm calculated coefficients, which were then compared with all gallery data to find the nearest neighbor. The proposed method reduces the computational time by eliminating one-to-one registration between the probe and gallery, which only calculates the distance between the parameters stored in the gallery and the probe. The comparison among different state-of-the-art methods shows that our method can be implemented in a real-time scenario. The challenge of applying ear images for human recognition is acquiring ear images. Although the proposed method is robust against natural occlusions, for effective recognition, it is obvious to capture the ear image with minimal occlusions, such as hair, headphones, etc. In our future work, we will analyze the method in more datasets and deploy it in edge devices for biometric applications.

Author Contributions: Conceptualization, M.M. and M.A.; methodology, M.M.; software, M.M.; validation, M.M., M.A. and P.H.-D.; formal analysis, M.M.; investigation, M.M.; resources, M.A.; data curation, M.M.; writing—original draft preparation, M.M.; writing—review and editing, M.M. and M.A.; visualization, M.M.; supervision, M.A.; project administration, M.A. and P.H.-D.; funding acquisition, M.A. and P.H.-D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are thankful to Syed Zulqarnain Gilani for his useful suggestions and feedback.

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

References

1. Zhao, W.; Chellappa, R.; Phillips, P.J.; Rosenfeld, A. Face recognition: A literature survey. *ACM Comput. Surv. (CSUR)* **2003**, *35*, 399–458. [\[CrossRef\]](#)
2. Maltoni, D.; Maio, D.; Jain, A.K.; Prabhakar, S. *Handbook of Fingerprint Recognition*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009.
3. Daugman, J. New methods in iris recognition. *IEEE Trans. Syst. Man Cybern. Part B (Cybernetics)* **2007**, *37*, 1167–1175. [\[CrossRef\]](#)
4. Lu, G.; Zhang, D.; Wang, K. Palmprint recognition using eigenpalms features. *Pattern Recognit. Lett.* **2003**, *24*, 1463–1467. [\[CrossRef\]](#)
5. Chen, H.; Bhanu, B. Contour matching for 3D ear recognition. In Proceedings of the 2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05), Breckenridge, CO, USA, 5–7 January 2005; Volume 1, pp. 123–128.
6. Ganapathi, I.I.; Ali, S.S.; Vu, N.S.; Prakash, S.; Werghi, N. A Survey of 3D Ear Recognition Techniques. *ACM Comput. Surv. (CSUR)* **2022**. [\[CrossRef\]](#)
7. Nejati, H.; Zhang, L.; Sim, T.; Martinez-Marroquin, E.; Dong, G. Wonder ears: Identification of Identical Twins from Ear Images. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 1201–1204.
8. Sharkas, M. Ear recognition with ensemble classifiers; A deep learning approach. *Multimed. Tools Appl.* **2022**, 1–27. [\[CrossRef\]](#)
9. Zhu, Q.; Mu, Z. PointNet++ and Three Layers of Features Fusion for Occlusion Three-Dimensional Ear Recognition Based on One Sample per Person. *Symmetry* **2020**, *12*, 78. [\[CrossRef\]](#)
10. Ramos-Cooper, S.; Gomez-Nieto, E.; Camara-Chavez, G. VGGFace-Ear: An Extended Dataset for Unconstrained Ear Recognition. *Sensors* **2022**, *22*, 1752. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Lei, Y.; Qian, J.; Pan, D.; Xu, T. Research on Small Sample Dynamic Human Ear Recognition Based on Deep Learning. *Sensors* **2022**, *22*, 1718. [\[CrossRef\]](#)
12. Lei, J.; Zhou, J.; Abdel-Mottaleb, M. Gender classification using automatically detected and aligned 3D ear range data. In Proceedings of the 2013 International Conference on Biometrics (ICB), Madrid, Spain, 4–7 June 2013; pp. 1–7. [\[CrossRef\]](#)
13. Chen, H.; Bhanu, B.; Wang, R. Performance evaluation and prediction for 3D ear recognition. In *International Conference on Audio-and Video-Based Biometric Person Authentication*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 748–757.
14. Chen, H.; Bhanu, B. Efficient recognition of highly similar 3D objects in range images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *31*, 172–179. [\[CrossRef\]](#)
15. Chen, H.; Bhanu, B. Human ear recognition in 3D. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 718–737. [\[CrossRef\]](#)
16. Chen, L.; Mu, Z.; Nan, B.; Zhang, Y.; Yang, R. TDSIFT: A new descriptor for 2D and 3D ear recognition. In Proceedings of the Eighth International Conference on Graphic and Image Processing (ICGIP 2016), International Society for Optics and Photonics, Tokyo, Japan, 29–31 October 2016; Volume 10225, p. 102250C.
17. Ganapathi, I.I.; Prakash, S.; Dave, I.R.; Joshi, P.; Ali, S.S.; Shrivastava, A.M. Ear recognition in 3D using 2D curvilinear features. *IET Biom.* **2018**, *7*, 519–529. [\[CrossRef\]](#)
18. Ganapathi, I.I.; Prakash, S.; Ali, S.S. Secure Multimodal Access with 2D and 3D Ears. In *Machine Learning for Intelligent Multimedia Analytics*; Springer: Singapore, 2021; pp. 1–20.
19. Ganapathi, I.I.; Ali, S.S.; Prakash, S. Geometric statistics-based descriptor for 3D ear recognition. *Vis. Comput.* **2020**, *36*, 161–173. [\[CrossRef\]](#)
20. Ganapathi, I.I.; Ali, S.S.; Prakash, S. Multi-resolution Local Descriptor for 3D Ear Recognition. In Proceedings of the BIOSIG 2019—18th International Conference of the Biometrics Special Interest Group, Gesellschaft für Informatik eV, Darmstadt, Germany, 18–20 September 2019.
21. Islam, S.M.; Davies, R.; Bennamoun, M.; Owens, R.A.; Mian, A.S. Multibiometric human recognition using 3D ear and face features. *Pattern Recognit.* **2013**, *46*, 613–627. [\[CrossRef\]](#)
22. 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. [\[CrossRef\]](#)
23. Prakash, S.; Gupta, P. Human recognition using 3D ear images. *Neurocomputing* **2014**, *140*, 317–325. [\[CrossRef\]](#)
24. Sun, X.; Wang, G.; Wang, L.; Sun, H.; Wei, X. 3D ear recognition using local salience and principal manifold. *Graph. Model.* **2014**, *76*, 402–412. [\[CrossRef\]](#)
25. Sun, X.P.; Li, S.H.; Han, F.; Wei, X.P. 3D ear shape matching using joint α -entropy. *J. Comput. Sci. Technol.* **2015**, *30*, 565–577. [\[CrossRef\]](#)

26. Maity, S.; Abdel-Mottaleb, M. 3D ear segmentation and classification through indexing. *IEEE Trans. Inf. Forensics Secur.* **2014**, *10*, 423–435. [\[CrossRef\]](#)
27. Zeng, H.; Dong, J.Y.; Mu, Z.C.; Guo, Y. Ear recognition based on 3D keypoint matching. In Proceedings of the IEEE 10th International Conference on Signal Processing Proceedings, Beijing, China, 24–28 October 2010; pp. 1694–1697.
28. Zeng, H.; Zhang, R.; Mu, Z.; Wang, X. Local feature descriptor based rapid 3D ear recognition. In Proceedings of the 33rd Chinese Control Conference, Nanjing, China, 28–30 July 2014; pp. 4942–4945.
29. 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. [\[CrossRef\]](#)
30. Cadavid, S.; Abdel-Mottaleb, M. Human identification based on 3D ear models. In Proceedings of the 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems, Crystal City, DC, USA, 27–29 September 2007; pp. 1–6.
31. Cadavid, S.; Abdel-Mottaleb, M. 3-D ear modeling and recognition from video sequences using shape from shading. *IEEE Trans. Inf. Forensics Secur.* **2008**, *3*, 709–718. [\[CrossRef\]](#)
32. Mahoor, M.H.; Cadavid, S.; Abdel-Mottaleb, M. Multi-modal ear and face modeling and recognition. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 4137–4140.
33. Passalis, G.; Kakadiaris, I.A.; Theoharis, T.; Toderici, G.; Papaioannou, T. Towards fast 3D ear recognition for real-life biometric applications. In Proceedings of the 2007 IEEE Conference on Advanced Video and Signal Based Surveillance, Genoa, Italy, 5–7 September 2007; pp. 39–44.
34. Theoharis, T.; Passalis, G.; Toderici, G.; Kakadiaris, I.A. Unified 3D face and ear recognition using wavelets on geometry images. *Pattern Recognit.* **2008**, *41*, 796–804. [\[CrossRef\]](#)
35. De Tré, G.; De Mol, R.; Vandermeulen, D.; Claes, P.; Hermans, J.; Nielandt, J. Human centric recognition of 3D ear models. *Int. J. Comput. Intell. Syst.* **2016**, *9*, 296–310. [\[CrossRef\]](#)
36. De Tré, G.; Vandermeulen, D.; Hermans, J.; Claes, P.; Nielandt, J.; Bronselaer, A. Bipolar comparison of 3D ear models. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*; Springer: Cham, Switzerland, 2014; pp. 160–169.
37. Mursalin, M.; Islam, S.M.S. Deep Learning for 3D Ear Detection: A Complete Pipeline from Data Generation to Segmentation. *IEEE Access* **2021**, *9*, 164976–164985. [\[CrossRef\]](#)
38. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In Proceedings of the Advances in NEURAL Information Processing systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5099–5108.
39. Kingma, D.P.; Ba, J. Adam: A method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
40. Mursalin, M.; Islam, S.M.S. EpNet: A Deep Neural Network for Ear Detection in 3D Point Clouds. In *International Conference on Advanced Concepts for Intelligent Vision Systems*; Springer: Cham, Switzerland, 2020; pp. 15–26.
41. Mursalin, M.; Islam, S.M.S.; Gilani, S.Z. 3D Morphable Ear Model: A Complete Pipeline from Ear Segmentation to Statistical Modeling. In Proceedings of the 2021 Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia, 29 November–1 December 2021; pp. 1–6.
42. Amberg, B.; Romdhani, S.; Vetter, T. Optimal Step Nonrigid ICP Algorithms for Surface Registration. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8. [\[CrossRef\]](#)
43. Gilani, S.Z.; Mian, A.; Shafait, F.; Reid, I. Dense 3D face correspondence. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 1584–1598. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Yan, P.; Bowyer, K.W. Biometric recognition using 3D ear shape. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1297–1308. [\[CrossRef\]](#) [\[PubMed\]](#)
45. 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. [\[CrossRef\]](#)