Cho, Siu-Yeung (2013) 3D ear shape reconstruction and recognition for biometric applications. Signal, Image and Video Processing, 7 (4). pp. 609-618. ISSN 1863-1711

3D ear shape reconstruction and recognition for biometric applications

Siu-Yeung Cho

Abstract This paper presents a new method based on a generalized neural reflectance (GNR) model for enhancing ear recognition under variations in illumination. It is based on training a number of synthesis images of each ear taken at single lighting direction with a single view. The way of synthesizing images can be used to build training cases for each ear under different known illumination conditions from which ear recognition can be significantly improved. Our training algorithm assigns to recognize the ear by similarity measure on ear features extracting firstly by the principal component analysis method and then further processing by the Fisher's discriminant analysis to acquire lower-dimensional patterns. Experimental results conducted on our collected ear database show that lower error rates of individual and symmetry are achieved under different variations in lighting. The recognition performance of using our proposed GRN model significantly outperforms the performance that without using the proposed GNR model.

Keywords Ear recognition · 3D shape reconstruction · Principal component analysis · Fisher's discriminant analysis

1 Introduction

Biometrics is the science of identifying a person using their physiological or behavioral features. One of the most popular biometrics identifier for adult is fingerprint. Unfortunately, it is too small and unstable for new born babies and very young children since their fingerprints are still under development. The fingerprints of a child only become more stable at the age of six. Nevertheless, the new born and child do have the need for reliable biometrics with respect to registry for identity cards and even biometric passport. The objective of this research is to investigate the alternative, i.e. **human ear**, as a biometrics identifier for babies at the time of birth and until 6 years old. Compared with other biometrics, like fingerprint, palmprint or blood sample, the advantages of ear shape are the availability of unique feature capturing, the simplicity of data collection (using only digital cameras) and higher user acceptability (non-intrusive and cost effective as compared to iris and retinal scanning).

The most significant work on ear identification that initiated researches on ear recognition was made by Iannarelli [1] in 1989, where over 10,000 ears were captured randomly in California for his manual approach. He even examined fraternal and identical twins. All examined ears were found to be unique though identical twins were found to have similar but not identical ear structures. In order to test the robustness and variability of ear biometrics, ear symmetry was also investigated. In Yan and Bowyer's [2] experiments on 404 persons, the mirrored left or right ear was used for matching and around 90 % of people's right ear and left ear were found symmetric. But some people's left and right ears have completely different shapes. They suggested that symmetry-based ear recognition cannot be expected to be accurate. There are some obvious advantages of ear over face and other biometrical methods. For examples, the variability over time is more for the face than the ear. The facial/emotional expression does not affect ear shape. People would feel less comfortable while taking part in face images enrollment (people tend to care how they look on photographs). The contactless feature of ear biometrics is undoubtedly no need to touch any devices, and therefore, there are no problems with hygiene. Ear requires a smaller image size which may imply smaller

S.-Y. Cho (✉)
Division of Engineering, The University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo 315100, China
e-mail: david.cho@nottingham.edu.cn

computational load. Thus, despite the attractiveness of face biometrics (e.g., they are easily verifiable by non-experts), ear biometrics is a promising technique which have both reliable and robust features and are extractable from a distance.

Although many ear recognition techniques have been able to deliver promising results, the task of robust ear recognition remains very difficult [3]. Indeed, there are at least two major problems in the current approaches: the illumination variation problem and the pose variation problem. Either of these two problems may cause significant degradation in the performance of ear recognition systems. 3D ear recognition is one of the methods which is able to tackle the above problems. Most recently, Islam et al. [16] proposed using fast 3D local feature matching and fine matching via an Iterative Closest Point (ICP) algorithm to obtain a robust ear recognition. But it is required to use a relative bulky and expensive range imaging technique to acquire the 3D ear shape for recognition. In this paper, a generalized neural reflectance (GNR) model is proposed as an alternative to enhance ear recognition in a way of delivering more robust output. Our approach is model-based method which differs from the other methods in that a single view ear image is required to synthesis other ear images under different lighting conditions. This generalized model [4] is established by using a hybrid structure of two neural activation functions, i.e., sigmoid and radial basis functions. Based on this model design, the diffuse model's parameters would be generalized by the sigmoid function, whereas the other parameters, such as specular reflectivity, could be approximated by the radial basis function. The radial based function is selected because of its separable capability in ill-posed hypersurface structure. All components for real ear images are generalized by this model, and a set of synthesis ear images can then be rendered in different occasions of illuminations. In our study, we only make use of one image in one view to estimate its ear surface, and thus, a set of synthetic ear images under different illumination conditions (i.e. different light source directions) can be synthesized. Our method for handling lighting variability in ear images differs from [5] in a way that our model is able to synthesis a large image database. In the recognition stage, a set of most expressive features is generated by the PCA to compress each ear representation, and then, the FDA is further implemented to generate a set of the most discriminant features so that different classes of training data can be classified. The identity of a test image can then be measured by means of different kinds of similarity measure. Our recognition approach is performed by using our collected database with totally 85 persons with 170 ear images containing both left and right side of ear images for each person.

The paper is organized as follow: Sect. 2 briefly describes a generalized reflectance model by use of a hybrid structure of neural models and shows how to synthesize ear images for training under different illumination. Section 3 describes the ear features extraction by means of the PCA and the FDA, and also different similarity measures can be expressed for the ear recognition. Section 4 presents experimental results, and finally, the conclusion will be drawn by Sect. 5.

2 Generalized model for ear surface reconstruction

It was shown that a $m \times n$ ear image can be formed as a convex object in the image space $\Re^{m\times n}$ under arbitrary combinations of point or extended light sources similar to face images [6]. Assume the surface of a convex object contains Lambertian reflectance surface that reflect light in diffuse reflection. Suppose that the surface, represented by $z\,(x, y)$, depends on the systematic variation of image brightness with surface orientation, where z is the height field and x and y are the 2D pseudo-plane over the domain Ω of the image plane. The Lambetian reflectance model uses to represent a surface illuminated by a single point light source which is given as follows:

$$R_{\text{Lambertian}} = \max\left(\eta \mathbf{n}\mathbf{s}^{\mathrm{T}}, \mathbf{0}\right), \tag{1}$$

where $\max\left(\eta \mathbf{n}\mathbf{s}^{\mathrm{T}}, \mathbf{0}\right)$ sets to zero for all negative components. η is the composite albedo, $\mathbf{s} = \begin{pmatrix}\cos\tau\sin\sigma & \sin\tau\sin\sigma & \cos\sigma\end{pmatrix}$ is illuminate source direction, and τ and σ denote the tilt and slant angles, respectively. $\mathbf{N} \in \Re^{(m\times n)\times 3}$ is defined to be a matrix whose rows are given as the surface normal, $\mathbf{n}$ be represented as follows:

$$\mathbf{n}\,(x, y) = \left(\frac{-p(x,y)}{\sqrt{p^2(x,y)+q^2(x,y)+1}} \;\; \frac{-q(x,y)}{\sqrt{p^2(x,y)+q^2(x,y)+1}} \;\; \frac{1}{\sqrt{p^2(x,y)+q^2(x,y)+1}}\right), \tag{2}$$

where $p = \frac{\partial z}{\partial x}$ and $q = \frac{\partial z}{\partial y}$ are the surface gradients. An ideal Lambertian surface requires a known and distant light source according to this model. But in most practical cases, the surface does not often contain the Lambertian surface because the light source is often located at a finite distance and an unknown position.

Specular component occurs when the incident angle of the light source is equal to the reflected angle. This component is formed by two terms: the specular spike and the lobe. Healey and Binford [7] derived the specular model by simplifying the Torrance-Sparrow model [8], in which the Gaussian distribution was used to model the facet orientation function. More sophisticated model based on the geometrical optics approach was also presented as the specular reflectance model [9], such that,

$$R_{\text{Specular}} = \kappa_{\text{spec}} \frac{L_i dw_i}{\cos\theta_r} \exp\left(-\frac{\alpha^2}{2\beta^2}\right), \tag{3}$$

where κ_{spec} represents the fractions of incident energy determined by the Fresnel coefficients and the geometrical

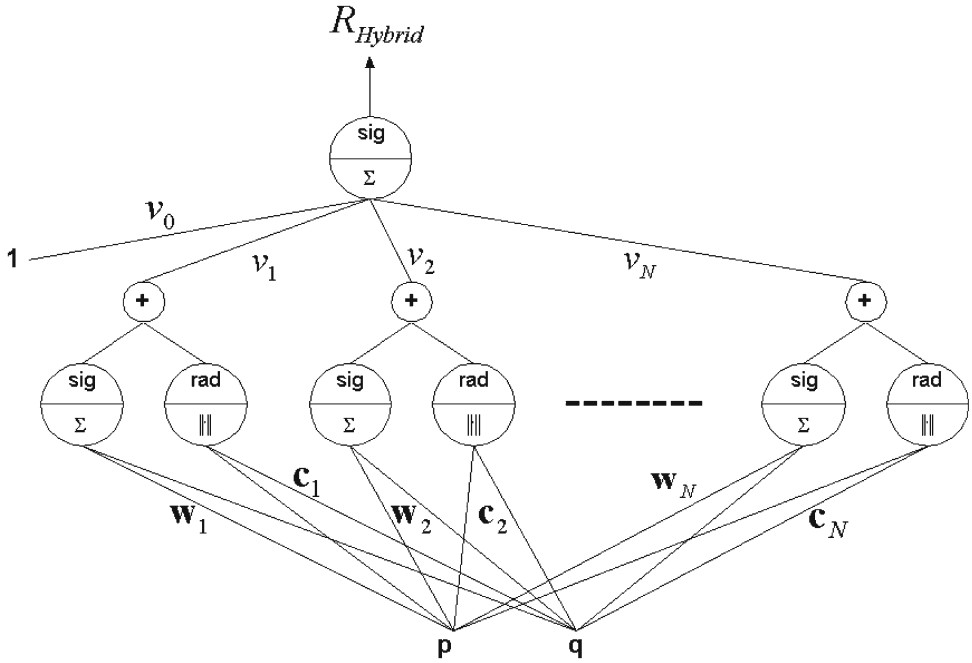

Fig. 1 Hybrid of sigmoid and radial basis activations for GNR model

attenuation factor. The term $\cos\theta_r$ describes the emitting angle that the radiance of the surface in the viewing direction is determined. As most object surfaces in the real world are neither purely Lambertian reflectance models, nor purely specular components, they are a linear combination of them. They are hybrid surfaces that include diffuse and specular components. Nayar et al. [9] formed a hybrid model to tackle the problem such that the model consists of three components: diffuse lobe, specular lobe and specular spike. In contrast, this paper describes a straightforward representation of the hybrid surface that the total intensity of the hybrid surface is the summation of the specular intensity and the Lambertian (diffuse) intensity as follow:

$$R_{\text{Hybrid}} = (1-\omega)\, R_{\text{Lambertian}} + \omega R_{\text{Specular}}, \tag{4}$$

where R_{Hybrid} is the total intensity for the hybrid surface, $R_{\text{Lambertian}}$ and R_{Specular} are the diffuse intensity and specular intensity, respectively, and ω is the weight of the specular component.

Although these models are widely used for the approximation of the reflectance components, the critical parameters (i.e., the light source and the viewing direction) are required a priori. Incorporating more reflectance parameters and effects is inevitable for generating a GNR model. In this paper, a neural-network self-learning scheme, based on the relationship between the surface orientation and the intensity, is exploited to model the unknown parameters for generalizing the reflectance model. Apparently, the use of a sigmoid activation model and a radial basis function model can provide approximations of the Lambertian model and the specular model, respectively, under the theoretical view of the universal approximation capability of neural networks [10,11]. These are clearly advantages of establishing a hybrid-type neural reflectance model [12], which combines the sigmoid and radial basis functions. The GNR model is shown in Fig. 1 and expressed as follows:

$$R_{\text{GNR}} = \varphi_{\text{sig}}\left(v_0 + \sum_{k=1}^{N} v_k \left(\varphi_{\text{sig}}\left(\mathbf{w}_k \mathbf{a}_{i,j}\right) + \varphi_{\text{rad}}\left(\left\|\mathbf{a}_{i,j} - \mathbf{c}_k\right\|\right)\right)\right), \tag{5}$$

where φ_{sig} and φ_{rad} are the sigmoid activation function and radial basis function, respectively. $\mathbf{w}_k$ is synapse weights of the sigmoid activation and $\mathbf{c}_k$ is centers of the radial basis function. The input vector $\mathbf{a}_{i,j} = \left(p_{i,j}\ q_{i,j}\right)^{\text{T}}$ acts a surface gradient vector in (i, j) coordinate of a face surface. The surface gradients would be optimized to form the optimal reflectance model $\hat{R}_{\text{GNR}}$ such that this model is equivalent to the given intensity image. This approach would enable us to generalize either the purely Lambertian surface or the non-Lambertian surfaces, which are most existing in the convex surface of the ear shape images. Using the above GNR model, the ear shape surface orientations can be reconstructed from the intensity image by solving a shape from shading (SFS) problem. In solving the SFS algorithm by the GNR model, the cost function is commonly used as follows:

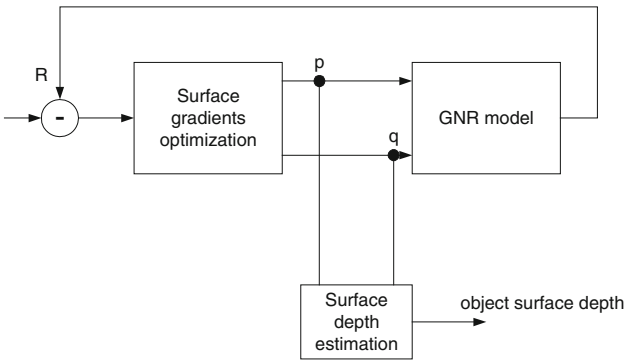

Fig. 2 The general learning framework for generating the object surface depth

$$E_{\mathrm{T}} = \iint_{\Omega} (I - R_{\mathrm{GNR}})^2 + \lambda \left(\left(\frac{\partial p}{\partial x} \right)^2 + \left(\frac{\partial p}{\partial y} \right)^2 + \left(\frac{\partial q}{\partial x} \right)^2 + \left(\frac{\partial q}{\partial y} \right)^2 \right) \mathrm{d}x\mathrm{d}y. \tag{6}$$

The first term is the intensity error term, and the second term is a smoothness constraint given by the spatial derivatives of p and q. λ is a scalar that assigns a positive smoothness parameters. Based on this objective function, the free parameters of the GNR model and the object surface gradients are determined by performing a unified learning mechanism. Through the learning process, the synapse weights, the radial basis centers as well as the surface depths are optimized by a specific learning framework which has been reported in [12,13]. The general learning framework is shown in Fig. 2. Throughout this learning framework, given an intensity image I, an error between the given intensity and the neural reflectance model can be computed and use to optimize the R_{GNR} parameters as well as the surface gradients by the cost function (6). The R_{GNR} parameters are optimized by a specific learning algorithm for the corresponding neural-network structure and the surface gradients are computed by a variational calculus approach on a discrete grid of points. Hence, the object surface depth can be estimated once the optimal surface gradients are being obtained. The details of this framework computation can be referred to the publication in [12] or [13].

The whole learning framework can normally converge within 10–20 iterations. Figure 3 demonstrates the reconstructed ear surfaces by the proposed GNR model. Figure 3a, b show the original single light source images for left and right ears, respectively, with the frontal view of 10 individuals from our collected database. The light source direction was chosen with 12° of the optical axis which the images do exhibit as little as shadowing. Figure 3c, b show the reconstructed ear surfaces by the GNR model for these corresponding individuals as shown in Fig. 3a, b, respectively. These ear surfaces can encode the corresponding surface normal fields which are synthesized ear images under arbitrary illumination conditions. We used the surface normal field obtained by the reconstructed surface to project the reflected intensities to the image plane by the arbitrary lighting directions for getting ear images under different illumination scenarios. Figure 4 demonstrates the samples of synthesized images of one person's ears (both left and right ears) in different lighting directions. Note that the synthesized ear images with the most extreme illumination scenario would not be used for training; however, those images would be used during testing for the recognition algorithm.

3 Extraction and recognition of ear features

In our study, about 40 synthesized images generated from each ear for each person were sufficient to provide for the different illuminate conditions. The next step is to extract the ear features from all these synthesized images to provide a representation for recognition. One of the simple ways is that the whole ear representation is projected down to a moderate-dimensional linear subspace in order to reduce the complexity and speed up the recognition process. Basically, the basis vector of this subspace, which is specific to ears, are commonly computed by performing principal component analysis (PCA) in which those basis vectors have been scaled by their corresponding eigenvalues. We then select the eigenvectors corresponding to the largest eigenvalues to be the basis vectors of the ears. We intended this as an approximation to finding the basis vectors by performing PCA directly on all the synthesized images of the ear under different illuminate conditions. In the simulations described in Sect. 4, the subspace of each ear had a dimension of 100 as this was good enough to specify all of the variability in the different illuminate conditions.

3.1 Principal component analysis (PCA)

Let an ear image $\mathbf{X}_i$ be a two-dimensional $m \times m$ array of intensity values. An image may also be considered as a vector of dimension m^2. Suppose that there are n ear images used for training $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n) \subset \Re^{m^2 \times n}$ and assumed that each image belongs to one of classes c. The covariance matrix is defined as follows:

$$\Psi = \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{X}_i - \overline{\mathbf{X}} \right) \left(\mathbf{X}_i - \overline{\mathbf{X}} \right)^{\mathrm{T}} = \Phi \Phi^{\mathrm{T}}, \tag{7}$$

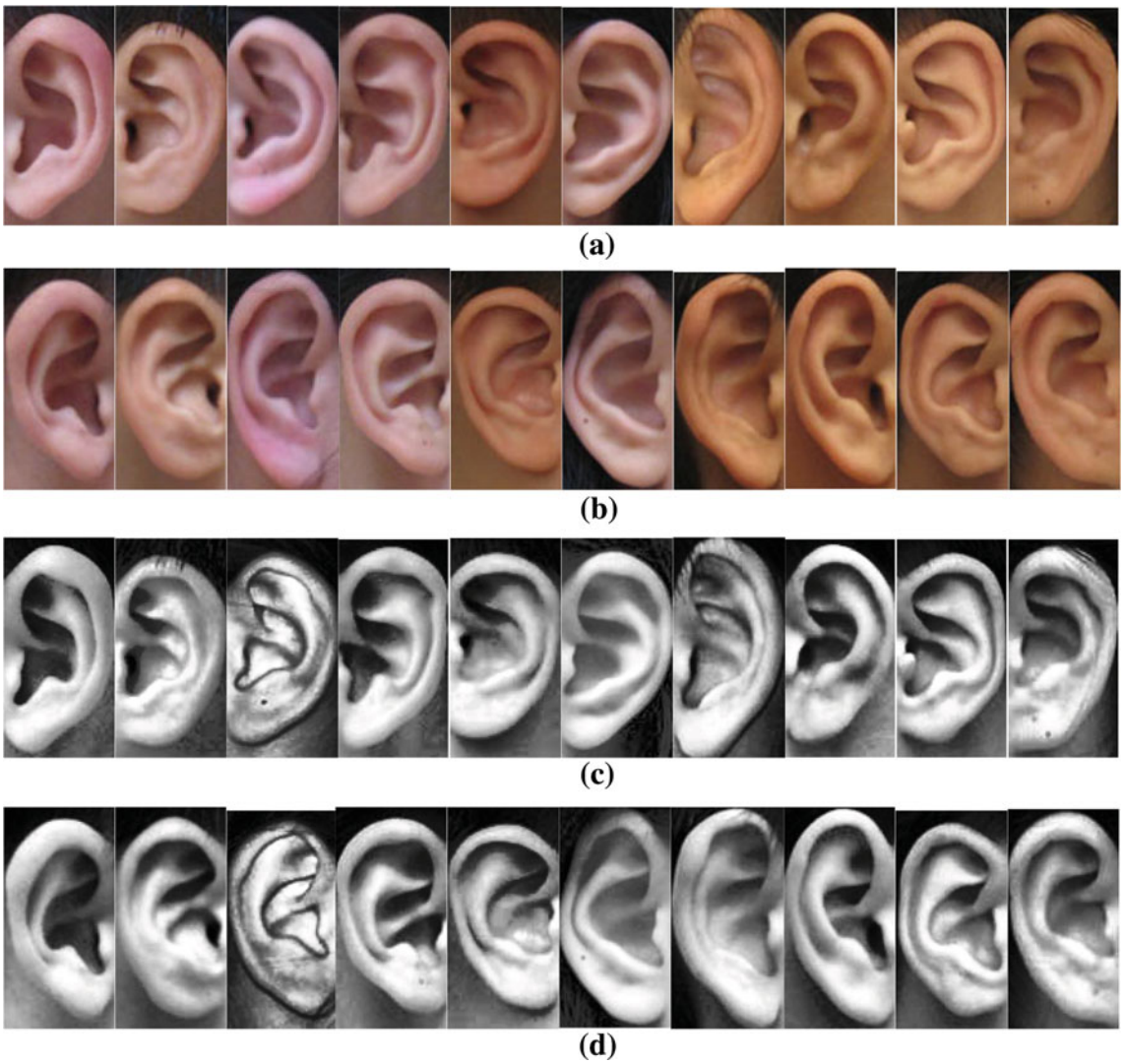

Fig. 3 The reconstruction results by the proposed GNR model: **a** *Left ear* images from 10 individuals; **b** *Right ear* images from 10 individuals (Note that all images are cropped and aligned by processing of ear localization which will be discussed later). **c** *Left ear* surfaces presented by the depth maps reconstructed by the proposed GNR model; **d** *Right ear* surfaces presented by the depth maps reconstructed by the proposed GNR model

where

$$\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i, \text{ and } \Phi = (\Phi_1, \Phi_2, \ldots, \Phi_n) \subset \Re^{m^2 \times n}. \quad (8)$$

The eigenvalues and eigenvectors of the covariance matrix Ψ are calculated. Let $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_r) \subset \Re^{m^2 \times r}$ where $r < n$ be the r eigenvectors corresponding to the r largest eigenvalues. Thus, for a set of original ear images $\mathbf{X}$, their corresponding eigen ear feature $\mathbf{Y} \subset \Re^{r \times n}$ can be obtained by projecting $\mathbf{X}$ into the linear sub-space (i.e. eigen-vector space) as follows:

$$\mathbf{Y} = \mathbf{A}^{\mathrm{T}}\mathbf{X}. \quad (9)$$

In the recognition process, a test ear image $\mathbf{X}_j$ is performed by first projecting $\mathbf{X}_j$ to $\mathbf{Y}_j$ at the eigen-vector space and then computing the metric to the eigen-vector representation of each ear $\mathbf{X}$ in the database. This metric is defined as a similarity measure to the closest projected eigen-vector space $\mathbf{Y}_j$ within $\mathbf{Y} \subset \Re^{r \times n}$. The face $\mathbf{X}_j$ is then assigned the identity of the closest representation.

However, the PCA paradigm does not provide any information for class discriminant. It means that the scatter being maximized is due not only to the between-class scatter that is useful for classification, but also to the within-class scatter that is unwanted information. Accordingly, the Fisher's discriminant analysis (FDA) is applied to the projection of the set of training samples in the eigen-vector space, and then,

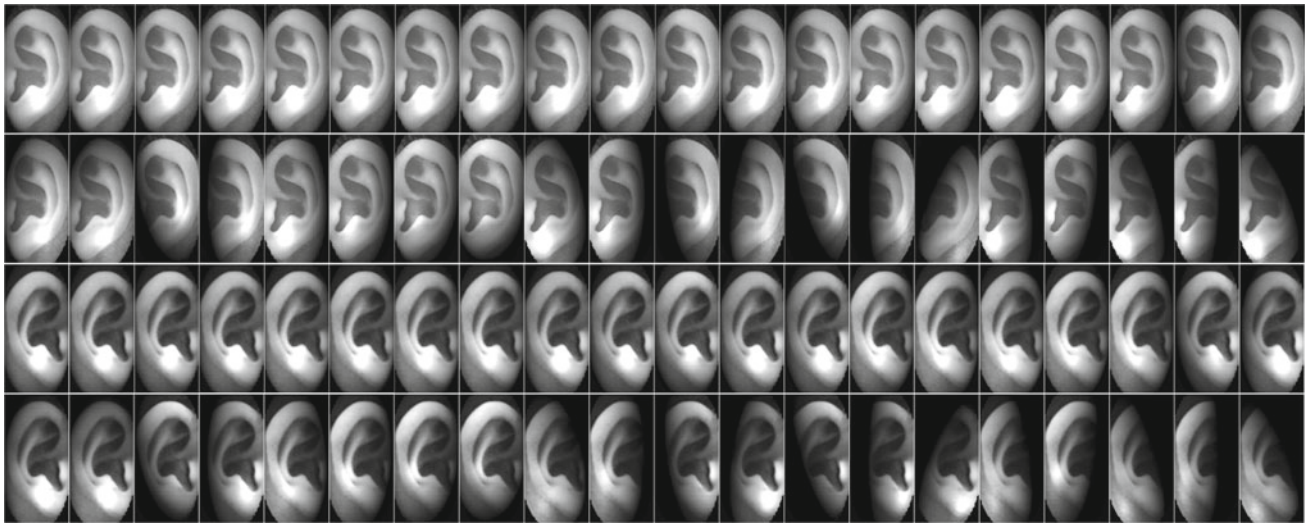

Fig. 4 Synthesized ear images (both *left* and *right*) of one individual under different illumination conditions

it finds an optimal subspace for classification in which the ratio of the between-class scatter and the within-class scatter is maximized [14,15].

3.2 Fisher's discriminant analysis (FDA)

The between-class scatter matrix is defined as

$$\mathbf{S}_B = \sum_{i=1}^{c} n_i \left(\overline{\mathbf{X}}_i - \overline{\mathbf{X}}\right)\left(\overline{\mathbf{X}}_i - \overline{\mathbf{X}}\right)^{\mathrm{T}}, \tag{10}$$

and the within-class scatter matrix is defined as

$$\mathbf{S}_W = \sum_{i=1}^{c} \sum_{\mathbf{X}_i \in n_i} \left(\mathbf{X}_i - \overline{\mathbf{X}}_i\right)\left(\mathbf{X}_i - \overline{\mathbf{X}}_i\right)^{\mathrm{T}}, \tag{11}$$

where $\overline{\mathbf{X}}_i$ is the mean image of class $\mathbf{X}_i$, and n_i is the number of samples in class $\mathbf{X}_i$. The optimal subspace, $\mathbf{Z}_{\mathrm{opt}}$ by the FDA is determined as follows [13]:

$$\begin{aligned}\mathbf{Z}_{\mathrm{opt}} &= \arg\max_{\mathbf{E}} \frac{\left|\mathbf{E}^{\mathrm{T}}\mathbf{S}_B\mathbf{E}\right|}{\left|\mathbf{E}^{\mathrm{T}}\mathbf{S}_W\mathbf{E}\right|}, \\ &= \left[\mathbf{z}_1\ \mathbf{z}_2\ \ldots\ \mathbf{z}_r\right]\end{aligned} \tag{12}$$

where $\{\mathbf{z}_i \,|\, i = 1, 2, \ldots, r\}$ is the set of generalized eigenvectors of $\mathbf{S}_B$ and $\mathbf{S}_W$ corresponding to the r largest generalized eigenvalues $\{\lambda_i \,|\, i = 1, 2, \ldots, r\}$, i.e.,

$$\mathbf{S}_B\mathbf{z}_i = \lambda_i \mathbf{S}_W \mathbf{z}_i, \quad i = 1, 2, \ldots, r \tag{13}$$

Note that there are at most $c - 1$ nonzero generalized eigenvalues, and so an upper bound on r is $c - 1$, where c is the number of classes.

In the recognition problem, it is difficult that the within-class scatter matrix is always singular. Thus, the rank: $\mathbf{S}_W \leq \min\{r, c\,(n_i - 1)\}$. In general, the value of r should be smaller than $n_i - c$. On the other hand, the rank: $\mathbf{S}_B \leq \min\{r, c - 1\}$, in which there are at most $c-1$ nonzero generalized eigenvectors. In other words, the FDA transforms the r-dimensional space into $(c - 1)$-dimensional space to classify c classes of ears. In order to overcome the complication of a singular $\mathbf{S}_W$, we propose an alternative feature extraction, which is achieved by using PCA to reduce the dimension of the feature space $N - c$ and then applying the standard FDA to reduce the dimension to $c - 1$. Thus, the feature vectors $\mathbf{f}_{\mathrm{query}}$ for any query ear images $\mathbf{X}_{\mathrm{query}}$ in the most discriminant sense can be calculated as follows:

$$\mathbf{f}_{\mathrm{query}} = \mathbf{Z}_{\mathrm{opt}}^{\mathrm{T}} \cdot \mathbf{A}^{\mathrm{T}} \cdot \mathbf{X}_{\mathrm{query}}. \tag{14}$$

Basically, it is noted that the FDA is a linear transformation which maximizes the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix of the projected samples. The results are globally optimal for linear separable data. Moreover, the separability criterion is not directly related to the classification accuracy in the output space.

3.3 Similarity measures

For recognition and classification purposes, we define the L_1, L_2 and cosine similarity measures and the nearest mean neighbor classification rule for ear recognition after obtaining

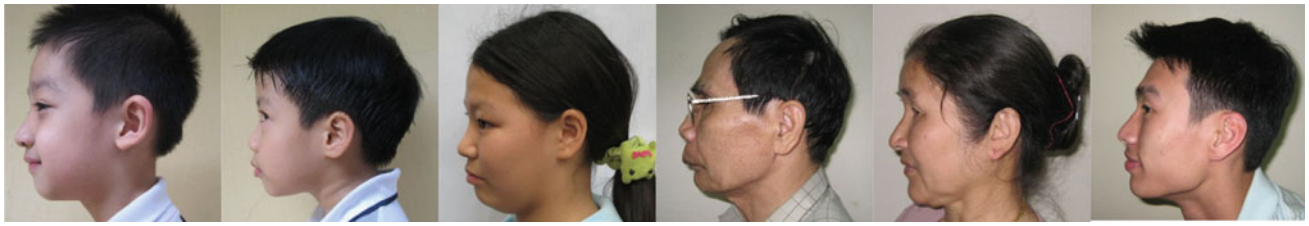

Fig. 5 Some examples of the collected profile image for our simulation

the feature vectors from the PCA/FDA paradigm. The nearest neighbor classification rule is defined as follows:

$$\sigma\left(\mathbf{f}, \mu_k\right) = \arg\min_j \sigma\left(\mathbf{f}, \mu_j\right) \rightarrow \mathbf{f} \in \omega_k. \tag{15}$$

The feature vector $\mathbf{f}$ is classified into the class of the closest mean μ_k based on the similarity measure σ. Similarity measures used in our studies include σ_{L_1}, σ_{L_2}, and $\sigma_{\cos}$, respectively, denote the L_1 distance measure, L_2 distance measure, and cosine similarity measure, which are defined as follows:

$$\sigma_{L_1}(\alpha, \beta) = \sum_i |\alpha_i - \beta_i|, \tag{16}$$

$$\sigma_{L_2}(\alpha, \beta) = (\alpha - \beta)^{\mathrm{T}}(\alpha - \beta), \tag{17}$$

$$\sigma_{\cos}(\alpha, \beta) = \frac{-\alpha^{\mathrm{T}}\beta}{\|\alpha\|\,\|\beta\|}, \tag{18}$$

where $\|\cdot\|$ denotes the norm operator.

4 Experimental results

We assess the feasibility and performance of our proposed recognition method performing on our collected ear image database. We have performed two experiments in this database. In the first one, tests were performed under variable illumination but fixed view and the goal was to compare the ear representation after reconstructing by the GNR model and without reconstructing by the GNR model. The second experiment was performed both left and right ears to evaluate the symmetrical and consistency of the proposed recognition approach. As demonstrated, our proposed recognition more effectively handles illumination variations for both ears.

4.1 Ear database

For this ear recognition simulation, we have collected ear images (both left and right ears) from 85 individuals in which they are distinguished by four different age groups (see Table 1 for the age distribution). For each individual, we collected 20 images for each ear in which we would select some samples randomly for training and the remaining samples for testing. All profile images were captured with relatively constant lighting (or captured with flash) and orthogonal viewpoint, and therefore, good condition of images can be obtained. Figure 5 shows some examples of the collected profile images for our simulation. After capturing the profile images, the processing of ear localization would be performed. This localization process was done by a semi-automatic processing, in which there are three functions that would be performed, they are as follows: rotating, cropping and reflecting the images. Human operation might first to crop the region of ear manually. The rotate function would use to rotate the ear image spatially to get all ear images the same orientation. The reflection function was provided for the need of comparing the symmetry of left ear and right ear. The operator may use the reflected left ear to match with right ear database. Figure 6 shows the interface of this ear localization process.

Table 1 Distribution of subjects' ages in the ear database

Age	Number of subjects
<9	7
9–18	18
19–50	47
>50	13
Total	85

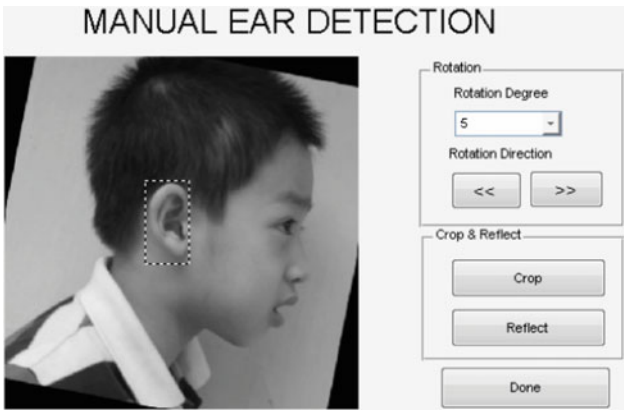

Fig. 6 Interface of the semi-automatic ear localization processing

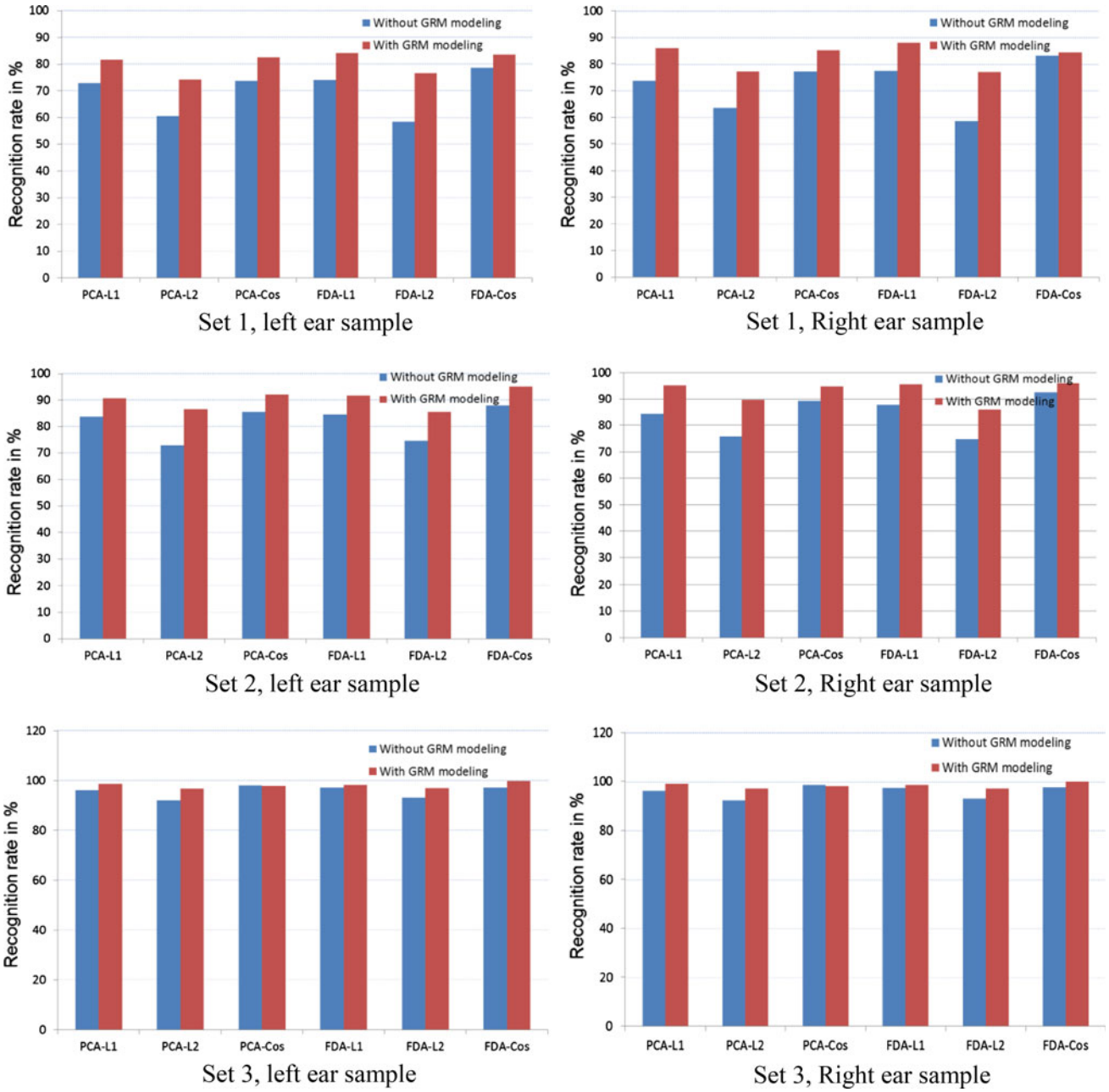

Fig. 7 Recognition rates of ear recognition for *left* and *right* ears individually. Each recognition method was trained with and without using GNR model to synthesis ear samples from 3D ear shapes. The testing simulations were conducted under different number of ear samples taken for training in which Set 1 contains one sample per ear, Set 2 contains 5 samples per ear and Set 3 contains 10 samples per ear

4.2 Experimental results of ear individual

As the previous mentioned, twenty images for each ear were collected such that total number of ear images for both left and right ears are 3,400 images for 85 individuals. In this experiment, three sets (Set A, B and C) of image samples for both left and right ears individually were divided for training in our experiments. In Set A, one image for each ear was selected as training and the remaining image samples were used as testing. In Set B, five images for each ear were selected as training and the remaining image samples were used as testing. In Set C, ten images for each ear were selected, and then, the other 10 images were used for testing. This experimental study was designed to compare the performance in between the same recognition method using the GNR model and without using the GNR model to reconstruct ear surfaces from different number of samples (1, 5 or 10 samples). As the ear images under different illumination scenarios were synthesized from the corresponding ear surfaces reconstructed by the GNR model, the training

Table 2 Recognition rates of ear symmetry by evaluating the mirrored left ear images

Extraction and similarity methods	PCA-L1		PCA-L2		PCA-Cos		FDA-L1		FDA-L2		FDA-Cos	
Synthesized by GNR model?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Set 1	65.6	73.5	54.5	66.8	66.2	74.3	66.7	75.8	52.5	69.0	70.7	75.3
Set 2	78.0	84.4	67.7	80.6	79.6	85.6	78.6	85.3	69.4	79.6	81.8	88.4
Set 3	83.5	85.7	80.0	84.1	85.3	85.0	84.4	85.4	80.9	84.3	84.4	86.8

Two extraction methods, principal component analysis (PCA) and Fisher's discriminant analysis (FDA), and three similarity measures, L1: σ_{L_1} metric in Eq. (16), L2: σ_{L_2} metric in Eq. (17), Cos: $\sigma_{\cos}$ metric in Eq. (18), are trained with and without using the GNR Model. The testing simulations were conducted under different number of ear samples taken for training in which Set 1 contains one sample per ear, Set 2 contains 5 samples per ear and Set 3 contains 10 samples per ear

Table 3 Recognition rates of ear symmetry by evaluating the mirrored right ear images

Extraction and similarity methods	PCA-L1		PCA-L2		PCA-Cos		FDA-L1		FDA-L2		FDA-Cos	
Synthesized by GNR model?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Set 1	66.2	77.4	57.1	69.5	69.5	76.6	69.6	79.2	52.7	69.3	74.8	75.9
Set 2	78.5	88.4	70.4	83.4	83.1	88.0	81.7	88.8	69.5	79.8	86.0	89.1
Set 3	83.6	86.1	80.3	84.4	85.6	85.2	84.7	85.8	80.9	84.3	84.8	86.9

Two extraction methods, principal component analysis (PCA) and Fisher's discriminant analysis (FDA), and three similarity measures, L1: σ_{L_1} metric in Eq. (16), L2: σ_{L_2} metric in Eq. (17), Cos: $\sigma_{\cos}$ metric in Eq. (18), are trained with and without using the GNR Model. The testing simulations were conducted under different number of ear samples taken for training in which Set 1 contains one sample per ear, Set 2 contains 5 samples per ear and Set 3 contains 10 samples per ear

size increased to around 40 images per sample in which apparently they represent all the illumination scenarios. The accuracy of recognition methods can then be dramatically improved by training these synthesis images even though the training size is rather greater than the original. Figure 7 shows the results from these simulations. Note that the ear recognition rates in very less samples taken (i.e., Set A) were about 79 and 81 %, respectively, corresponding to the PCA and the FDA representations by training the 40 synthesis images for each ear, whereas the recognition rates was about 70 % by only training the original images in Set A. The results support that whenever using this GNR model, good recognition rates are achieved by means of ear representation in low-dimensional subspaces in approximation of the different illumination conditions with less sample taken.

4.3 Experimental results of ear symmetry

In this experiment, we performed the evaluation of ear symmetry of recognition for left and right ears. In this evaluation, we produced a mirror ear database for testing in which a mirror left ear was created which is the reflection of the original right ear. Thus, a mirror right ear was created from the reflected left ear. The purpose of this evaluation is to test whether the system does still perform well if there is only one ear (left or right) is able to be captured or one ear image (left or right) was corrupted. This would be an application for forensic analysis. After creating the mirror ear database for left and right ears, the model would recognize those mirrored ear images individually, i.e., the model trained by left ear samples would recognize the mirrored left ear and vice versa. The results of mirrored left ear and mirrored right ear are shown in Tables 2 and 3, respectively. Again, the results support that whenever using this GNR model, good recognition rates are achieved by means of ear representation in low-dimensional subspaces in approximation of the different illumination conditions even though the tested ears are created and reflected by another side of the ears.

5 Conclusion

In this paper, we addressed a problem in ear recognition by variations in lighting. Therefore, we presented a framework of ear recognition which requires a small number of images of an ear in several fixed postures and illuminated by a single point light source at unknown positions to generate a rich representation of the ear images which are useful for recognition. The main idea of our method is to make use of the proposed GNR model to transform the given ear into reconstructed 2.5D ear surfaces. Using all these reconstructed ear surfaces, a full set of ear images can then be synthesized under different illumination conditions. These synthesized images can be used for training in many cases for variation of lighting. The recognition method, in this paper, simply uses the PCA to reduce the dimensionality of ear representation and

further uses the FDA to enhance the classification to discriminate the features. The experimental results demonstrate the performance of ear recognition is markedly improved after using the proposed GNR reconstruction and transformation. We believe that our method can be applied to practical cases of ear recognition under variations in illuminations, although the PCA/FDA may not be ideal solution for huge number of ear samples. In fact, we can use other more advance recognition methods instead of PCA/FDA to achieve results practically, such as using Support Vector Machine as a classifier, which may be able to achieve better results than PCA/FDA. We also believe that this method is applicable to other object recognitions in industrial applications where similar representations are used.

References

1. Iannarelli, A.: Ear Identification. Paramount Publishing Company, Fremont (1989)
2. Yan, P., Bowyer, K.W.: Biometric recognition using 3D ear shape. IEEE Trans. PAMI **29**(8), 1297–1307 (2007)
3. Burge, M., Burger W.: Ear biometrics in computer vision. In: 15th International Conference of Pattern Recognition, ICPR 2000, pp. 826–830 (2000)
4. Cho, S.-Y., Chow, T.W.S.: Robust face recognition By using generalised neural reflectance model. Neural Comput. Appl. J. **15**, 170–182 (2006)
5. Chen, H., Bhanu, B.: Human ear recognition in 3D. IEEE Trans. PAMI **29**(4), 718–737 (2007)
6. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 643–660 (2001)
7. Healy, G., Binford, T.O.: Local shape from specularity. Comput. Vis. Graph. Image Process. **42**, 62–86 (1988)
8. Torrance, K.E., Sparrow, E.M.: Theory for off-specular reflection from roughened surfaces. J. Opt. Soc. Am. **57**, 1105–1114 (1967)
9. Nayar, S.K., Ikeuchi, K., Kanade, T.: Determining shape and reflectance of hybrid surfaces by photometric sampling. IEEE Trans. Robotics Autom. **6**, 418–430 (1990)
10. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximation. Neural Netw. **2**, 359–366 (1989)
11. Park, J., Sandberg, I.W.: Universal approximation using radial basis function networks. Neural Comput. **3**, 261–257 (1991)
12. Cho, S.Y., Chow, T.W.S.: Enhanced 3D shape recovery using the neural-based hybrid reflectance model. Neural Comput. **13**, 2617–2637 (2001)
13. Cho, S.Y., Chow, T.W.S.: Neural computation approach for developing a 3-D shape reconstruction model. IEEE Trans. Neural Netw. **12**(5), 1204–1214 (2002)
14. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)
15. Swets, D.L., Weng, J.: 'Using discriminant eigenfeatures for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell. **18**, 831–836 (1996)
16. Islam, S.M.S., Davies, R., Ben-namoun, M., Mian, A.S.: Efficient detection and recognition of 3D ears. Int. J. Comput. Vis. **95**, 52–73 (2011)