Pattern Recognition ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Automated human identification using ear imaging

Ajay Kumar*, Chenye Wu

Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 3 September 2010
Received in revised form
5 May 2011
Accepted 24 June 2011

Keywords:
Biometrics
Ear identification
Personal identification
Ear segmentation

ABSTRACT

This paper investigates a new approach for the automated human identification using 2D ear imaging. We present a completely automated approach for the robust segmentation of curved region of interest using morphological operators and Fourier descriptors. We also investigate new feature extraction approach for ear identification using localized orientation information and also examine local gray-level phase information using complex Gabor filters. Our investigation develops a computationally attractive and effective alternative to characterize the automatically segmented ear images using a pair of log-Gabor filters. The experimental results achieve average rank-one recognition accuracy of 96.27% and 95.93%, respectively, on the publicly available database of 125 and 221 subjects. Our experimental results from the authentication experiments and false positive identification verses false negative identification also suggest the superiority of the proposed approach over the other popular feature extraction approach considered in this work.

1. Introduction

Reliability in personal authentication is the key issue for the stringent security requirements in many application domains ranging from airport surveillance to electronic banking. Many physiological characteristics of humans, i.e., biometrics, are typically invariant over time, easy to acquire, and unique to each individual. Therefore the biometrics traits are increasingly adopted for civilian applications and no longer confined for forensic identification. Most of the current research in biometrics is focused on face, fingerprint, gait, signature, iris, palmprint, or hand-geometry [11]. However, there have been little efforts to investigate the human ear for personal authentication despite its significant role in forensic science. The ear is quite attractive biometric candidate mainly due to its (i) rich and stable structure that is preserved since birth and is quite unique in individuals, (ii) being invariable to the changes in pose and facial expression, and (iii) relatively immune from anxiety, privacy, and hygiene problems with several other biometric candidates. Therefore automated personal identification using ear images has been increasingly studied for possible commercial applications.

1.1. Related work

Human ear has attracted several studies on its individuality and uniqueness. Iannarelli [1] has manually measured the distance between predicted points on human ear. He has extensively examined 10,000 ears and concluded on their uniqueness. The utility of Iannarelli's approach for the automated ear recognition is quite limited since this approach required accurate estimation of a stable reference points for measurements, which is very difficult in real environment. The ear image characterization using Voronoi diagram is illustrated in [2,3]. However, the work in Ref. [2] is largely conceptual and lacks experimental results on any ear database. Hurley et al. [4] have presented one of the most promising approaches for the automated ear identification and developed a new method of localizing ear shape features using force field transformation. Authors have employed the database of 63 users to illustrate the appearance-based ear recognition. Chang et al. [7] investigated principal component analysis to characterize ear and face using *eigenear* and *eigenface*, respectively. The rigorous experimental results detailed in Ref. [7] achieve competing performance from these two independent modalities and illustrate that the combination of these two modalities can achieve convincing improvement in the performance for the human identification. Shape of human ear is known to be quite complex and has also been characterized using geometrical parameters for the biometric identification. Mu et al. [15] have attempted such characterization of ear shapes from the gradient of ear images, using long axis based shape and structural feature extraction, and achieved promised results on the database of 77 subjects. Moreno et al. [23] have detailed the usage of neural network classifiers on ear profile images derived from the gray level ear images. Ear shape features [19,20] have shown to offer promising performance but on small size databases. Nanni and Lumini [30] have recently exploited the color information from the ear images and detailed the selection of color spaces, using the sequential forward floating selection, when the fitness

* Correspondence to: PQ 835, Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. Tel.: +852 2766 7254; fax: +852 2774 0842.
E-mail address: ajaykr@ieee.org (A. Kumar).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.06.005

function is related to the optimization of performance from the ear recognition.

Human ear is highly curved 3D surface and therefore provides rich 3D discriminent features for the biometric identification. Bhanu and Chan [10] have exploited the local surface shape descriptors for the 3D ear identification. These authors have also suggested iterative closest point matching (ICP) in Ref. [11] for the matching of 3D ear shapes. One of the most promising approaches for the automated ear identification using 3D imaging is presented by Yan and Bowyer [18]. This approach is based on the active contour function to localize ear shape and uses ICP that uses k–d tree data structure to efficiently search for closest point. The use of ICP for characterization of 3D ear shape has also been detailed in Refs. [31,32]. The slow acquisition speed of 3D ear imaging (such as Vivid 910 3D digitizer employed in the literature), limits the online usage of 3D imaging for the civilian applications. The currently employed 3D digitizer in the literature is also quite expensive and large in size. Therefore our focus in this work has been to exploit the 2D ear images that can be conveniently acquired from conventional digital camera.

1.2. Our work

This paper presents a completely automated approach for the personal identification using 2D ear imaging. Automated and accurate segmentation of exact region of interest, from the acquired gray level ear images, is quite difficult. However the success of the feature extractor and the achievable accuracy/performance significantly depends on the accuracy for such automated ear segmentation. Therefore significant efforts are required to develop robust techniques for the completely automated ear segmentation. We present a new approach for the automated ear segmentation which has been proven to be quite effective in achieving robust segmentation of the curved region of interest. Our ear segmentation approach effectively utilizes combinations of morphological operations and Fourier descriptors to achieve reliable ear segmentation. Second set of efforts in this paper are directed to ascertain the performance for the automated ear identification using new feature extraction approaches on the ear images. In this context, we explore the orientation features using even Gabor filters and also ascertain the effectiveness of localized gray level phase information using complex Gabor filters. We finally present a computationally attractive and promising alternative to characterize the automatically segmented ear images using a pair of log-Gabor filters. This paper also presents comparative performance evaluation from the proposed feature extraction approach with those using promising alternatives presented in the literature. The rigorous experimental results presented in this paper, from the ear image dataset of 125 subjects and 221 subjects, suggest its superiority over the alternative approaches presented in the literature. The ear image database developed in this work is made publicly available to foster further research efforts in the automated ear identification.

The rest of this paper is organized as follows. Section 2 describes our completed automated approach for the segmentation of ear shape images. This section is followed by the details of the feature extraction approach using a pair of log-Gabor filters in Section 3. This section also details the two other feature extraction approaches for characterizing the gray level orientation information and the local phase information from the automatically segmented ear images. Section 4 details the experiments and the corresponding results from the proposed and prior approaches. This is followed by discussion in Section 5 and the key conclusions from this paper are summarized in Section 6.

2. Automated ear segmentation

One of the most challenging aspects of the 2D automated ear identification is related to the automated and accurate segmentation of ear or the region of interest that encloses discriminant gray level information. The human ear is highly curved 3D surface and therefore generates in uneven reflections which also generate shadows. Therefore the acquired images have uneven illumination, low contrast, and often surrounded by hairs and skin with varying pigmentation. Therefore the most challenging aspects of developing automated ear identification, as observed from our study, is to develop a robust algorithm that can reliably segment a fixed region of interest for the extraction of features that are more stable in a given class/subject. The block diagram of our automated ear segmentation method is shown in Fig. 1.

Each of the acquired images is firstly subjected to the preprocessing that consists of smoothing with a Gaussian filter which helps to suppress noise, followed by histogram equalization. The resulting image is used to automatically generate a binarized mask that can outline the *surrounding* region of interest. This step requires the binarization of image using Otsu's threshold. The resulting image however generates the masks of varying sizes (primary due to uneven and varying illumination) and therefore this thresholding limit is automatically adjusted (Fig. 1) until the mask area is less than the predefined limit.

The pre-processed image is also simultaneously used to ascertain the shape mask that reliably captures the ear shape. This is quite challenging due to the presence of shadows, surrounding skin, hair and our efforts to localize the ear shape using spatial and spatial-domain filtering yielded poor results (extracted regions were not stable). We therefore pursued with a series of gray scale morphological operations to extract the ear shape.

The morphological operations typically compare the ear image with another known object, i.e., structuring element. The shape and size of such structuring element is determined during the training phase from the prior knowledge of acquired ear images with the objective of effective ear localization (within a given region)

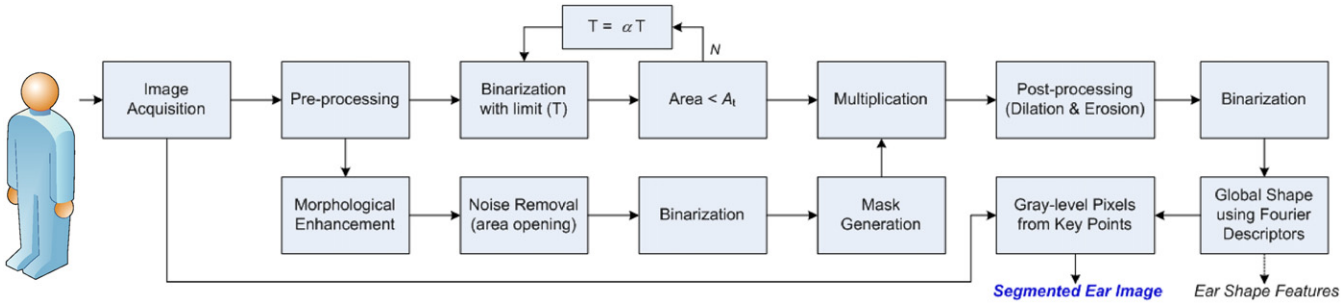

Fig. 1. Block diagram of the proposed method for the automated ear segmentation.

Please cite this article as: A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.06.005

and noise elimination. The shape of structuring element is primarily selected to probe the presence of wide line-like shape corresponding to the width which is determined by the size of structuring element [36]. The initial morphological operations employed in our approach can be represented as follows:

$$w(x,y) = g(x,y) - (g(x,y) \ominus s) \odot s \quad (1)$$

where $\ominus$ represents gray-scale dilation and $\odot$ represents gray-scale erosion by structuring element s, $g(x, y)$ is the acquired image after histogram equalization and $w(x, y)$ is the resulting output image. The closing operation, i.e., image dilation followed by the erosion, helps to smooth the ear shape contours, breaks narrow isthmuses, and eliminates small islands and occasional sharp peaks in $g(x, y)$. The *black top-hat* operation resulting from Eq. (1) has been quite effective to highlight objects in the complex background. The ear shape contours are enhanced and more effectively connected after this black top-hat operation. In summary, series of employed morphological operations help to significantly accentuate the structural information corresponding to the ear shape in the acquired images and this can also be observed from the sample images shown in Figs. 2 and 3.

The resulting image is further subjected to gray scale area opening operations to remove the noise. This gray scale image is binarized to extract the ear shape boundaries and subjected to

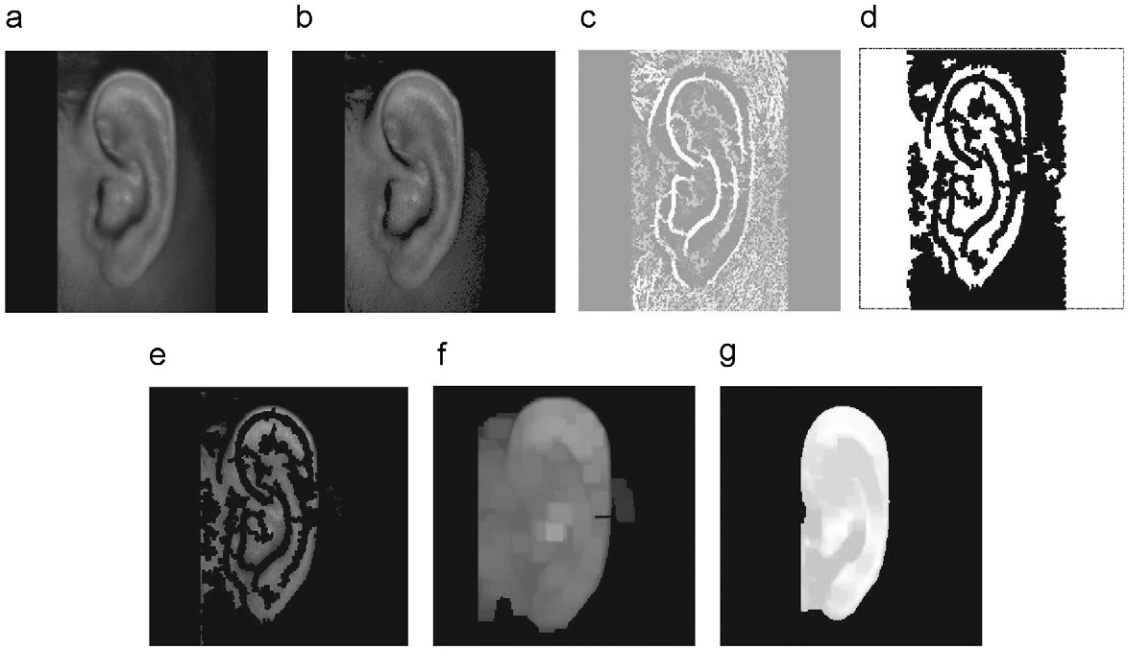

Fig. 2. Automated extraction of ear shape from the acquired images; (a) acquired image, (b) image after histogram equalization, (c) images after closing and opening, (d) generated mask, (e) image after multiplication of mask, (f) image after dilation and (g) ear shape image after noise elimination.

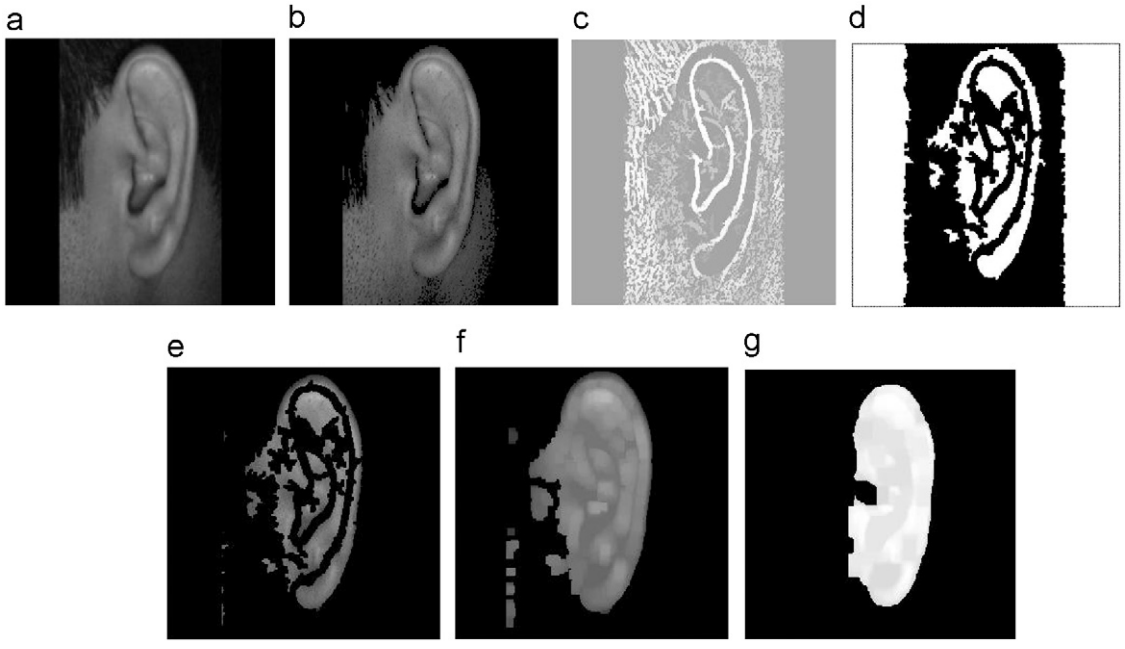

Fig. 3. Automated extraction of ear shape from the acquired images; (a) acquired image, (b) image after histogram equalization, (c) images after closing and opening, (d) generated mask, (e) image after multiplication of mask, (f) image after dilation and (g) ear shape image after noise elimination.

Please cite this article as: A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.06.005

mask generation. The key task here is to effectively remove the skin region surrounding the ear shape boundaries. The resulting mask is combined with the ear silhouette generated earlier to completely eliminate the areas surrounding the ear.

2.1. Extraction of ear shape using Fourier descriptors

The boundary tracing of ear shape image (Figs. 2 and 3(g)) is then employed to generate the ear shape boundary. These boundaries generated using the employed approach is shown in Fig. 4 for three subjects. Such exterior boundaries generated are often not smooth as can also be observed from Fig. 4(a–c), and therefore further processing is necessary. In our approach, we employed Fourier descriptors and this process is briefly discussed in the following. Let us assume that the boundary tracing of the ear shape image generates N point contour in x-y plane which can be represented by (x_0, y_0), (x_1, y_1), (x_2, y_2), … (x_{N-1}, y_{N-1}) with an arbitrary starting point (x_0, y_0) [38]. Thus the extracted ear shape boundary can be represented as sequence of coordinate pairs, i.e., $c(k)=[x(k), y(k)]$. In this representation, the x- and y-axis can be considered as real and imaginary axis, respectively, to convert ear contour into a complex number, $c(k)=x(k)+iy(k)]$. The discrete Fourier transform (DFT) of such ear contour $c(k)$ is obtained as follows:

$$b(z) = \sum_{k=0}^{N-1} c(k) e^{-i2\pi zk/N} \quad \text{for } z = 0, 1, \ldots N-1 \tag{2}$$

The above equation generates N complex coefficients which represents Fourier descriptors of the extracted ear shape contour. These Fourier descriptors $b(z)$ directly reflect the changes in the curvature of ear shape contour. In order to extract the global ear shape and eliminate the accompanying contour noise, we reconstruct the ear shape

with only first L coefficients, i.e., $b(z)=0$ for $\forall z > L-1$. The approximated ear shape contour is thus computed as follows:

$$\tilde{c} = \frac{1}{L} \sum_{z=0}^{z=L-1} b(z) e^{i2\pi zk/N} \tag{3}$$

The reconstructed ear shape contour for the three different subjects is shown in Fig. 4 (using $L=30$, as employed for all the experimental results reported in this paper).

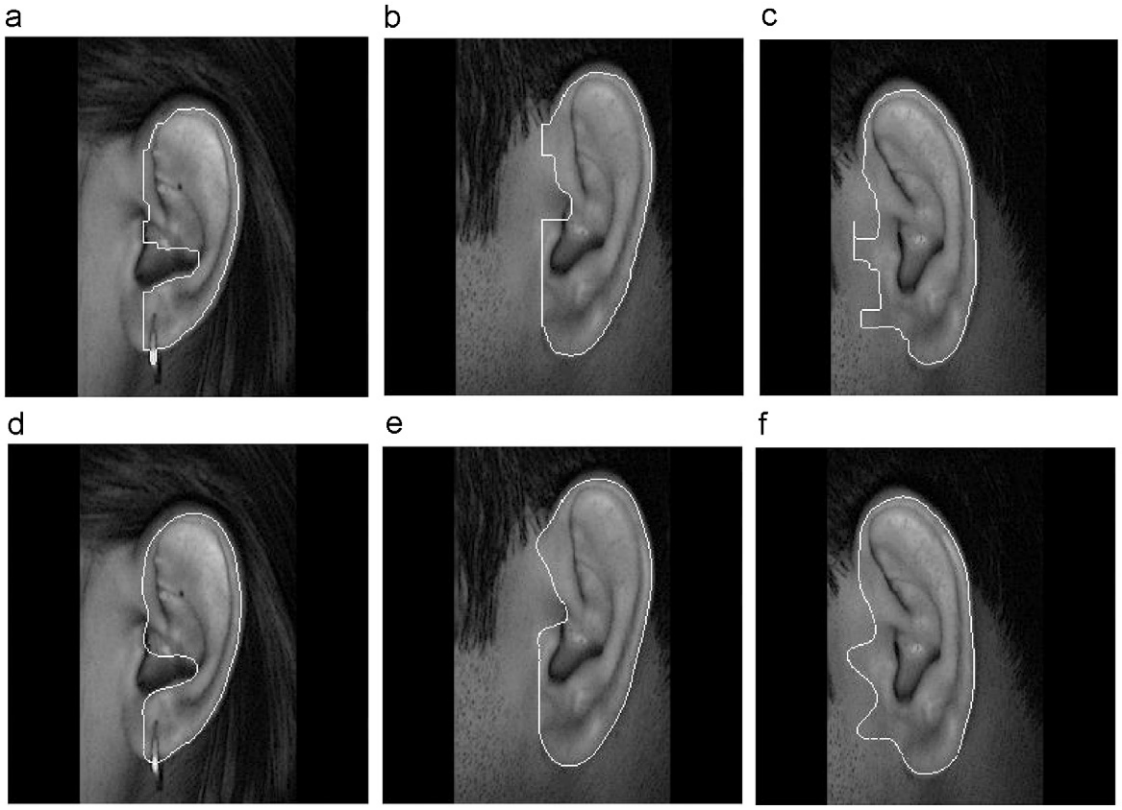

Fig. 4. Automated localization of ear shape boundaries before using Fourier descriptors in (a)–(c) and after using Fourier descriptors in (d)–(f).

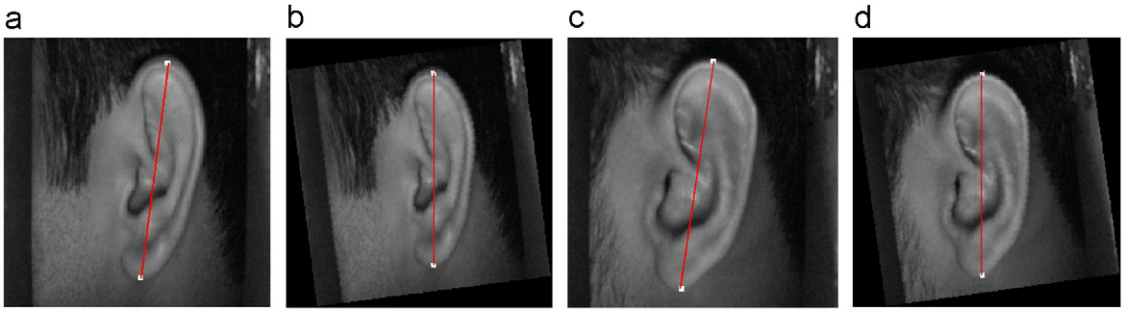

Fig. 5. Extraction of stable key points before the normalization in (a) and (c) and after normalization in (b) and (d).

Please cite this article as: A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.06.005

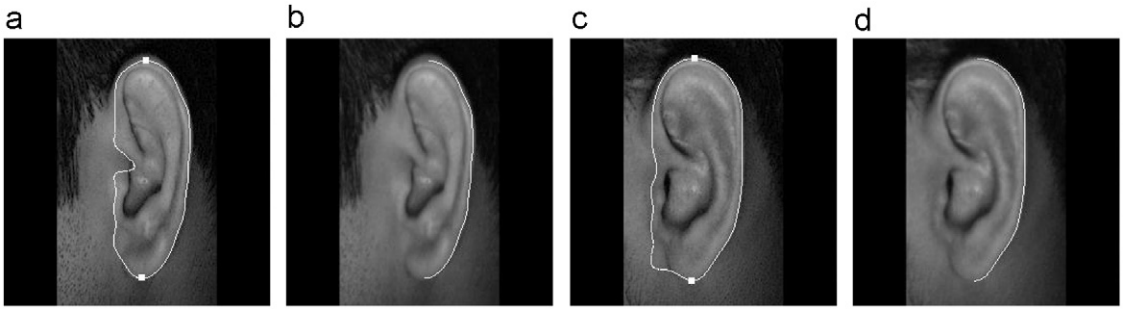

Fig. 6. The clockwise region from the normalized ear images (a), (c) is used to extract stable gray level region (b), (d) for feature extraction.

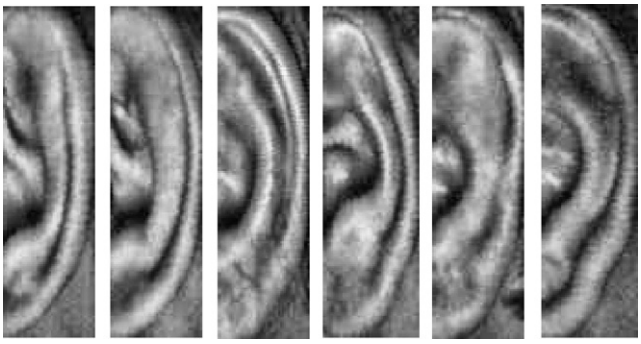

Fig. 7. Normalized and enhanced ear image samples from six subjects.

The two key points on this contour which achieve the maximum distance between them are selected as reference points on the reconstructed ear shape contour. Fig. 5 reproduces some images, along with their rotation, to illustrate that the extraction of two reference points using employed approach is robust. The region of interest from these reference points is automatically extracted and represents the most stable set of gray level pixels in the clockwise region (Fig. 6). Fig. 7 illustrates set of such automatically normalized, segmented and enhanced ear images employed for the feature extraction and classification.

3. Feature extraction and matching

The accuracy of automated ear identification approach is highly influenced by the nature of extracted features and the employed matching process. In this work, we investigated three new feature extraction and matching approaches for the identification of automatically segmented ear images. The extraction of phase information using 1D log-Gabor and 2D Gabor filters was firstly investigated for the ear identification. The spatial orientation of gray level shape features can be efficiently encoded using even Gabor filters and such orientation details can be employed to generate templates for the matching. Therefore such feature extraction approach using a bank of even Gabor filter was also investigated and achieved promising results as detailed in Section 4. The complex Gabor filters can also be employed to extract local phase information, as commonly used in the iris recognition [37], and was also investigated in this work. In the following section we briefly describe the feature extraction and corresponding matching approaches investigated in this work.

3.1. Phase encoding using multiscale log-Gabor filters

Automated identification of localized gray level information requires simultaneous measurements in spatial and spatial-frequency domain. Gabor filters have invited lot of applications in texture analysis and biometrics measurements, mainly due to their orientation selectivity, and maximum possible joint resolution in spatial and spatial-frequency domain. However, the Gabor filter-based characterization of images is known to under emphasize low frequency contents and over emphasize on the high frequency contents in a given image. Field [12] proposed log-Gabor filters to overcome the bandwidth constraints in traditional Gabor filters. These log-Gabor filters always have null dc component and desirable high-pass characteristics, i.e., the frequency response of log-Gabor filters have elongated tail in the high frequency part which helps to efficiently capture fine details from the high-frequency areas of a given image.

The frequency response of log-Gabor filter in frequency domain is defined as

$$G(f) = \exp\left(\frac{-(\log(f/f_o))^2}{2(\log(\sigma_f/f_o))^2}\right) \tag{4}$$

with f_0 is the central frequency and σ_f is the scaling factor of the radial bandwidth B. Fig. 8 illustrates the spatial domain representation of log-Gabor filters employed for the extraction of unique features from the normalized and enhanced ear images in this work. The radial bandwidth in octaves from such filters can be expressed as follows:

$$B = 2\sqrt{2/\ln 2} * |\ln(\sigma_f/f_0|) \tag{5}$$

Fig. 9 illustrates the amplitude spectrum of typical log-Gabor filter on log scale. This spectrum falls of at $1/f$ rate, which is similar to the natural image spectrum, and therefore the log-Gabor filters are well adapted for encoding information from the natural images.

The feature encoding approach employed in this work uses a pair of log-Gabor filters and is illustrated in Fig. 10. Each of the pixel rows is firstly unwrapped to generate complex log-Gabor filter response. The resulting convolved form of the signal is complex valued. We then apply phase quantization [34] to extract binary phase templates which represents encoded phase information. Thus the feature extraction process encodes each of the enhanced ear images into a pair (or complex) of binary templates corresponding to number of bits of information, in a manner similar to as illustrated in Eq. (13). The matching scores between two ear images are generated using the normalized Hamming distance between their corresponding binarized templates. The Hamming distance D_{PQ} between two $Y \times Z$ size complex bitwise ear template, P and Q, is computed as follows:

$$D_{PQ} = \frac{\sum_{y=1}^{Y}\sum_{z=1}^{Z}\{P_r(y,z) \oplus Q_r(y,z) + P_i(y,z) \oplus Q_i(y,z)\}}{2 \times Y \times Z} \tag{6}$$

where $\oplus$ represents bitwise XOR operation, P_r and Q_r are the real part of the templates while P_i and Q_i are the imaginary part of the bitwise ear templates P and Q, respectively. The bit-wise shifting of ear templates, i.e. left and right, is employed during the matching

Please cite this article as: A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.06.005

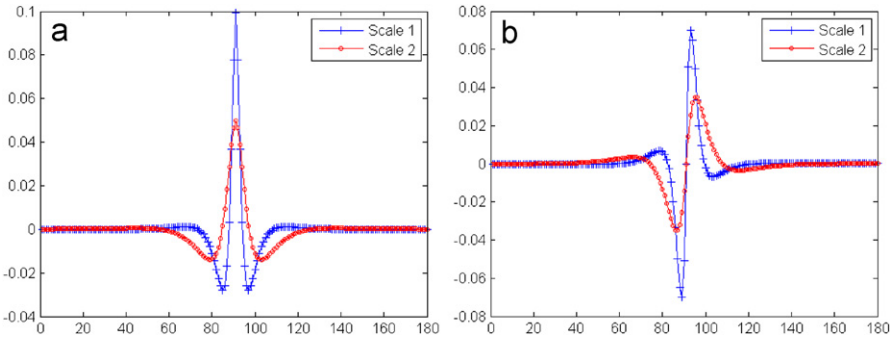

Fig. 8. The (a) real and (b) imaginary Log-Gabor filters in spatial domain having bandwidth of 2 octaves and a center frequency of 1/18.

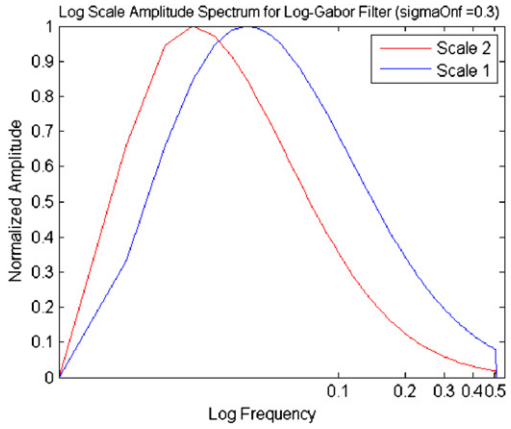

Fig. 9. Amplitude spectrum of the employed log-Gabor filter on log sale.

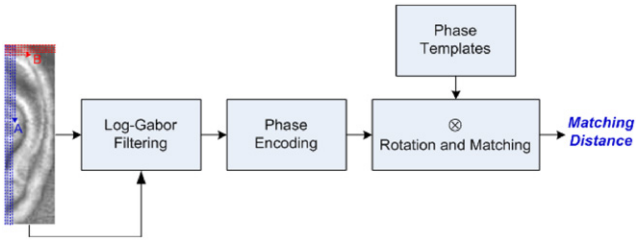

Fig. 10. Feature extraction and matching from the normalized images using a pair of log-Gabor filters.

as it helps to account for the translational errors in during the image localization. The center wavelength of 18 and the σ_f/f_0 ratio of 0.55 was empirically fixed for the log-Gabor filters employed in this work. The selection of these parameters was performed during the training phase, by only using training data, and fixed for all the experiments in this paper. It is quite reasonable to expect that the variations in these parameters can degrade the performance, as also illustrated in reference [37] for the iris recognition. In our experiments, the best of the Hamming distance generated from the template shift of 10 bits (5 bits left and right each) are used as the final matching distance.

We also extracted the feature templates when the 1D signals are generated from the column scanning (Fig. 10) are employed. These two orthogonal directions will encode the different gray level spatial information which is likely to be least correlated. We therefore generate two sets of templates, referred to as log-Gabor(A) and log-Gabor(B) in the experiments. The Hamming distances from these two respective templates are combined to generate consolidated matching distance. In this work, we investigated two new approaches for consolidating these two Hamming distances, i.e., weighted minimum and weighted product combination, apart from the popular weighted score combination [27] method for the score level combination. The consolidated matching scores $\tilde{s}$ from the two orthogonal templates are generated as follows:

$$\text{Weighted minimum}: \tilde{s} = \min_{\forall s_a, s_b}(w * s_a,\ s_b), \quad 1 \le w \le 2 \tag{7}$$

$$\text{Weighted product}: \tilde{s} = s_a * (s_b)^{w-1} \quad \forall s_a,\ s_b, \quad 1 \le w \le 2 \tag{8}$$

$$\text{Weighted sum}: \tilde{s} = s_a * (w-1) + s_b * (2-w), \quad \forall s_a, s_b, \quad 1 \le w \le 2 \tag{9}$$

where the s_a and s_b, respectively, represents the corresponding Hamming distances obtained from the two templates in the orthogonal directions (Fig. 9). The weight w is empirically estimated using only the training data during the training phase. The consolidated matching score $\tilde{s}$ is used to ascertain the identity of the unknown ear. In case of verification task, this consolidated matching score is compared with a decision threshold and the user is subsequently assigned into genuine or impostor class.

3.2. Orientation encoding using even Gabor filters

The spatial orientation details of the curved ear shape can be encoded using a bank of even Gabor filters. Such encoding scheme based on orientation of textured features has shown to provide some of the best promising results for palmprint [28] and also for the finger knuckle [29]. Therefore we also investigated such an approach using bank of even Gabor filters which achieved promising results. Each of the segmented and enhanced ear images are firstly convolved with a set of self similar Gabor filters. The even Gabor filter is a product of Gaussian and sinusoid denoted as follows:

$$G_{\theta_n}(x,y) = \Gamma\, e^{-\pi(x^2+y^2/\sigma^2)} \cos(2\pi f(x\cos\theta_n + y\sin\theta_n)) \tag{10}$$

where Γ represents the magnitude of the Gaussian envelope, σ corresponds to the standard deviation of Gaussian function, θ_n represents the orientation of the sinusoidal wave, and f is the frequency of the sinusoid. A set of self similar even Gabor filters is employed to ascertain the localized orientation of gray levels. These self similar filters are obtained by the rotation of even Gabor filter G' in Ω intervals, i.e. from 0, π/Ω, $2\pi/\Omega$ to $(\Omega-1)\pi/\Omega$. The multiple responses from the Ω even Gabor filters are then compared to select a prominent orientation at every pixel in the normalized ear image as follows:

$$T(x,y) = \arg\{\max_{\forall n = 1,2,..\Omega}\{G'_{\theta_n}(x,y) \otimes E(x,y)\}\} \tag{11}$$

where $G'_{\theta_n}(x,y)$ represents zero mean $G_{\theta_n}(x,y)$, i.e., obtained by subtracting the mean values, $\otimes$ represent convolution operation,

Please cite this article as: A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.06.005

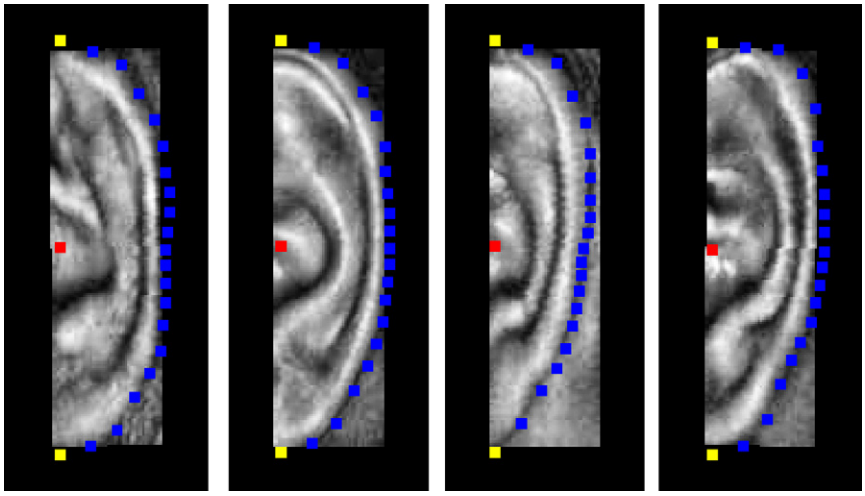

Fig. 11. Location of distance points for the automatically extracted shape features from the four typical subjects ear images in our database.

and $E(x, y)$ is the normalized and enhanced ear image. Since we considered only six orientations, the orientation index (features) are binary encoded using *three bits* and stored as representative feature for every pixel corresponding to $E(x, y)$. The matching of two feature maps from two ear images, i.e. T and Q corresponding to the template and query ear images respectively, is ascertained from the angular distance $D(T,Q)$ as follows:

$$D(T,Q) = \frac{\sum_{y=1}^{Y}\sum_{z=1}^{Z}\sum_{k=0}^{3} T_k^b \oplus Q_k^b}{3 \times Y \times Z} \tag{12}$$

where T_k^b and Q_k^b is the kth bit plane of $Y \times Z$ template T and Q, respectively. The possible translational changes in the segmented ear image are accounted by generating multiple matching scores, with the translated version of the template in horizontal and vertical directions, and employing the minimum of these matching scores as the final matching score. We also investigated to use weighted combination for the different orientation, instead of maximum as in Eq. (11), and evaluated the possible improvement in the performance.

3.3. Phase encoding using complex Gabor filter

The Gabor filter based phase encoding of localized gray level information has shown to offer excellent results for texture segmentation, fingerprint identification, and iris recognition [13]. Therefore the Gabor filter response from the even and the odd Gabor filters were also employed in this work to ascertain the performance. The Gabor filtered output is demodulated to extract the phase information which is quantized to four levels for each possible quadrant in complex plane using the convolution response from the normalized ear image. The encoding of the phase information from the complex Gabor filters $H(\theta,F,\sigma)$ into the feature map $F(x, y)$ is achieved as follows:

$$\begin{aligned} F(x,y)_r &= 1 \quad \text{if} \quad Re\{H(\theta,f,\sigma) \otimes E(x,y)\} \geq 0 \\ F(x,y)_r &= 0 \quad \text{if} \quad Re\{H(\theta,f,\sigma) \otimes E(x,y)\} < 0 \\ F(x,y)_i &= 1 \quad \text{if} \quad Im\{H(\theta,f,\sigma) \otimes E(x,y)\} \geq 0 \\ F(x,y)_i &= 0 \quad \text{if} \quad Im\{H(\theta,f,\sigma) \otimes E(x,y)\} < 0 \end{aligned} \tag{13}$$

Each of the normalized ear images is used to generate binarized feature map using Eq. (13). The normalized Hamming distance between the query and the template feature map is computed in a manner similar to as detailed in Eq. (6). In order to account for the further possible translation in the normalized ear images, bitwise shifting of ear templates, five pixels left and five pixels right, is employed to generate the best possible matching distance.

3.3.1. Shape features

We also extracted geometrical features that can describe the shape information from the automatically extracted ear shape images (Fig. 1). The distance between the two reference points was used for the normalization and we sampled N points every $180/N$ degrees. The experiments were performed for the various N values and $N=20$, which uses 19 $(N-1)$ distances as features, achieves best results. Euclidean distance between the shape features is used to compute the matching distance for the ear identification using shape features. Fig. 11 illustrates the location of these sampled distances on the shape images from four subjects in our database.

4. Experiments and results

The experimental results in this paper are presented on publicly available dataset of 465 ear images from the 125 subjects. The left ear images from 125 subjects/volunteers were acquired using simple imaging setup that employed a digital camera in an indoor environment. The images were acquired over a period of 9 months in indoor environment and no additional illumination was employed. Each of the volunteers sat on a chair and a digital camera was fixed to acquire the region of interest which ensured the presence of ear in imaging window. All the subjects in the database are in the age group 14–58 years and provided at least 3 images. Some images from this dataset are reproduced in Fig. 12 and the images in this dataset have significant scale, translational and rotational variations. A Gaussian filter with $\mu=20$ and $\sigma=5$ was employed in the preprocessing stage. The predefined limits for the binarization, i.e., area A_t and α in Fig. 1, is fixed at 25,000 and 0.95, respectively, for all the experiments in this paper. A squared structuring element of size two pixels was selected for the closing operation in Eq. (2). The resulting images were subjected to area opening, i.e., removal of isolated and connected pixels if their count is less than 80 while considering their connectivity in 8-directions (box shape). Each of the acquired images are automatically segmented and normalized, using the approach detailed in Section 2, to 50×180 pixel size region of interest images. In addition to this database, we also formed a larger dataset that included all the images from the database in Ref. [14,22] and this larger dataset is also made available in Ref. [33]. Therefore a larger version of dataset had 753 images from 221 subjects and is also employed for the performance evaluation. The automated extraction of the field lines from the ear images using force field transform [5] approach, which was also implemented by us for the performance comparison, on the sample

Please cite this article as: A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.06.005

Fig. 12. Typical image samples from our acquired database from the four subjects.

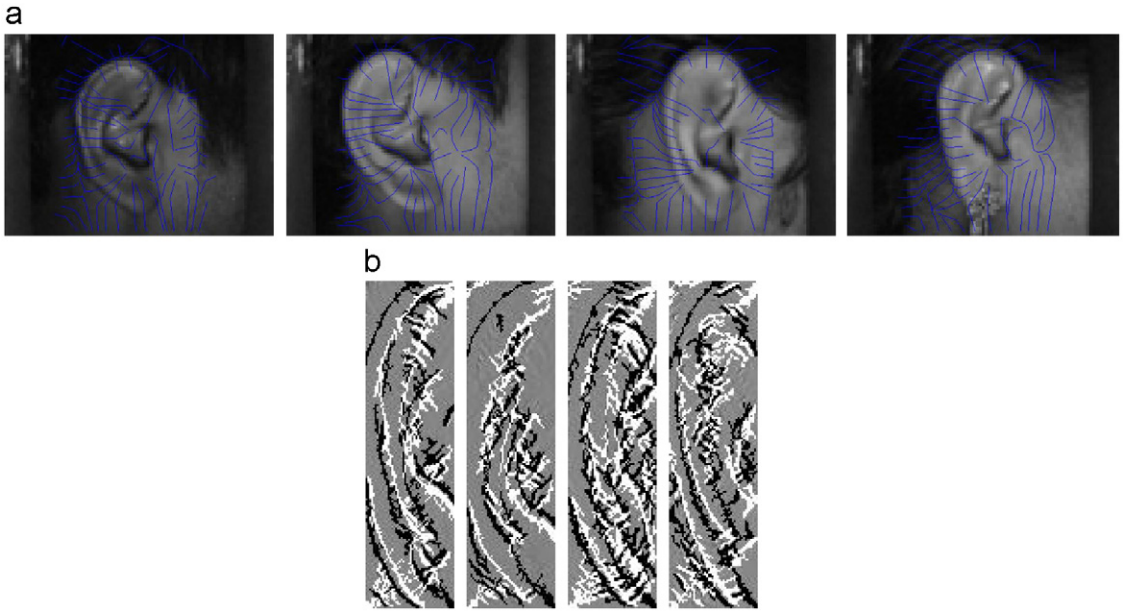

Fig. 13. (a) Extracted field lines for the force field transform approach and (b) thresholded convergence maps for the four typical images in our dataset.

images in our database is shown in Fig. 13. This figure also illustrates the convergence field on the ear images from four different subjects in our database.

The automatically segmented and enhanced ear images were employed to extract the orientation features using even Gabor filters. The gray level representation of these orientation features

Please cite this article as: A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.06.005

on the images from the four subjects are shown in Fig. 14. The gray level representation of phase information extracted using complex Gabor filters, as detailed in Section 3.3, is shown in Fig. 15 for the four sample images. As detailed in Section 3, each of the automatically normalized ear images were subjected to the feature extraction in two orthogonal directions using log-Gabor filters. Fig. 16 illustrates the gray level representation of such phase information generated from the log-Gabor filters, in each of the two orthogonal directions, for three different subjects in our dataset.

In this paper, we employed one image from every subject as test image while the remaining two images were employed as training images. Such matching was repeated for three times for each of the three possible independent test images and the average of the experimental results are presented in this paper. Each of the training images were further rotated, left and right by 5 degrees, and the best matching score was utilized as the matching distance from the corresponding training image. The cumulative match characteristics from this set of experiments, each for the 125 and 221 subject dataset, are illustrated in Fig. 17. This figure comparatively illustrates the performance from our implementation of force field transform, eigen ear (PCA), shape features, Gabor orientation, Gabor phase, and using the suggested log-Gabor filter based approach. It can be ascertained from this figure that the proposed approach using log-Gabor filters achieved superior/best performance while the shape features and eigen ear approach have performed poorly on our dataset.

Table 1 illustrates the average rank one recognition rate from the proposed approach for the comparative evaluation. The rotation of training dataset achieved significant performance improvement and the rank one recognition rate; for example the rank one recognition rate without employing the rotation for log-Gabor(A), log-Gabor(B), weighted sum, weighted minimum from the were 93.33%, 88%, 94.13%, and 94.13%, respectively, for the 125 subject dataset (Table 1). The weight parameters for the equation (9), (7), (8) corresponding to the weighted sum, weighted minimum, and weighted product combination results shown in Fig. 17 and Table 1 were (1.87, 1.12, 1.5), (1.9, 1.05, 2), respectively, for the 221 subject and 125 subject database. The weights for the three orientations for consolidating matching scores (Table 1) were empirically fixed as (0.3, 0.3, 0.4) and (0.4, 0.5, 0.1), respectively, for the 221 and 125 subject dataset during the experiments. One of the important observations in Table 1 is related to the performance from the weighted Gabor orientation features. The proposed approach using weighted orientation has been quite successful in improving the performance, as compared to those when weights are not employed. The average rank-one recognition results from three combination schemes using log-Gabor filters shown in Table 1 suggest that the these are quite close/competing but the product (weighted product) for 125

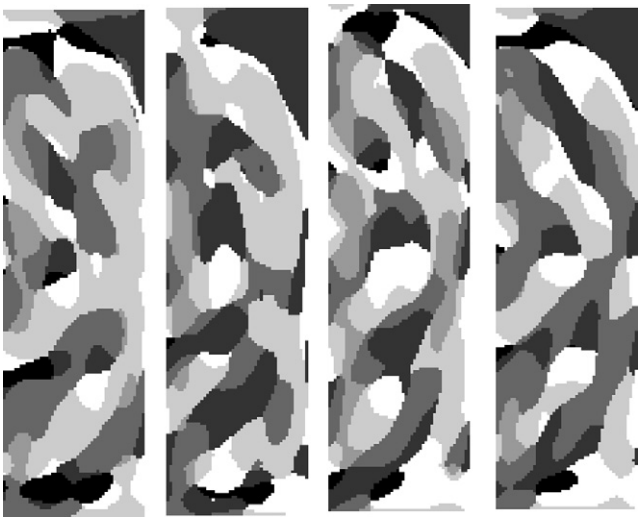

Fig. 14. The gray level representation of encoded orientation features using even Gabor filters from the four typical images in our dataset.

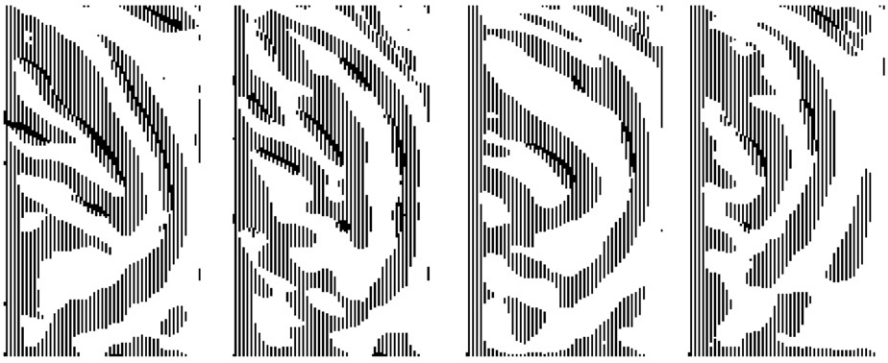

Fig. 15. The gray-level representation of phase features using complex Gabor filters from the four typical images in our dataset.

Fig. 16. The gray level representation of encoded phase features using a pair of log-Gabor filters on the three typical normalized images in our dataset.

Please cite this article as: A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.06.005

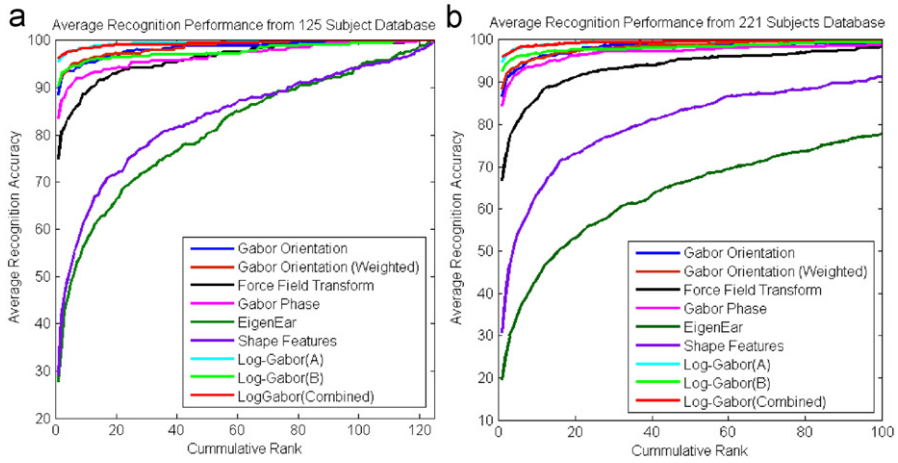

Fig. 17. The cumulative match characteristics for the average recognition performance on the dataset with (a) 125 subjects and (b) 221 subjects.

Table 1
Average rank one recognition accuracy from experiments.

Feature extraction and matching		Database I (125 subjects) (%)	Database II (221 subjects) (%)
Log-Gabor	Log-Gabor(A)	95.4667	94.7210
	Log-Gabor(B)	90.4000	92.6094
	Weighted sum	95.4667	95.93
	Weighted min	95.4667	95.63
	Weighted product	96.27	95.78
Gabor orientation	Original	88.533	86.727
	Weighted	90.4	88.3861
Gabor phase		83.4667	84.4646
Shape feature		29.0667	30.6184
Force field transform [5]		74.9333	66.6667

and weighted sum for the 221 subject dataset achieves best performance.

Although the key objective of our work was to develop promising online ear recognition approach, we also performed the experiments for the verification tasks using the same train/test protocol and the same parameters as employed for the recognition task. The receiver operating characteristics (ROC) from the experiments on the two dataset is shown in Fig. 18. The ROC in Fig. 18 illustrates the superiority of the proposed approach using log-Gabor filter and the combination of pair of log-Gabor filter responses has been more effective for the authentication problem as improvements shown in ROC are quite distinct and consistent. The equal error rates from the authentication experiments from the log-Gabor(A), log-Gabor(B), Gabor orientation, weighted Gabor orientation, Gabor phase, force field transform, were (3.14%, 5.01%, 5.46%, 5.32%, 6.95%, 13.75%) and (4.51%, 7.7%, 7.56%, 6.62%, 9.62%, 13.7%), respectively, for the 221 and 125 subject dataset. The corresponding equal error rate from the combination of two log-Gabor filters using weighted sum, weighted minimum, and weighted product combination were (3.01%, 2.85%, 3.0%) and (3.85%, 3.97%, 3.36%), respectively, for the 221 and 125 subject dataset. In summary our experimental results from the authentication experiments also suggest superior performance using the proposed approach using log-Gabor filters.

We also performed several experiments to ascertain the sensitivity of the proposed method to the selection of log-Gabor filter parameters. These experimental results are illustrated in Fig. 19 where results from the test dataset for the variation in the center wavelength and the ratio σ_f/f_0 are presented. The results from Fig. 19 (a) achieve best case average equal error rate of 4.31% from the test data (for 125 subject dataset) if we are allowed to select the better values of the center wavelength from the observation of best possible performance on test data. The observed results in figure suggest that the experimental results are quite sensitive to the selection of log-Gabor filter parameters.

An automated biometric system should also be able to efficiently recognize unknown subjects, i.e., able to reject those subjects not enrolled in the database. In order to ascertain such capability explicitly, we selected last 20 subjects from the two ear databases (125 and 221 subjects) as unknown subjects. These unknown subjects are then identified from the proposed identification approach to ascertain the performance. Fig. 20 shows the plot of number of unknown subjects identified as unknown verses known subject rejected as unknown. These experimental results also suggest that the log-Gabor based combination scheme outperforms other approaches considered in this work. The performance from the proposed ear identification for the FPIR (false positive identification rate) and FNIR (false negative identification rate) was also observed on the two databases. We selected last 20 subjects as unknown subjects from the test data and rest of the subjects, were used as test data (subjects 1–105, 1–201) for the performance evaluation. Fig. 21 illustrates the performance for the FPIR VS FNIR characteristics for the two databases. The performance improvement using the combination of log-Gabor filters is observed to be quite consistent as compared to those from other approaches considered in this work. In summary, the proposed scheme using log-Gabor filters is also expected to improve FPIR and FNIR performance for their usage in the automated ear identification.

5. Discussion

The accuracy of automated ear identification highly depends on the accuracy of segmenting the ear or the region of interest. In this context, Ref. [21] has also shown that the rank one recognition rate improves from 62% to 96% when the automated ear segmentation is replaced by more reliable manual ear segmentation. One of the possible approaches for the automated ear detection is to employ Haar-like features and train the classifier using ear and non-ear samples (Adaboost detector). Such approach can however be useful for the ear detection but not for the exact localization of ear shape boundaries for which some heuristic approach, similar to the one developed and evaluated in this work, should be employed. The automated method ear

Please cite this article as: A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.06.005

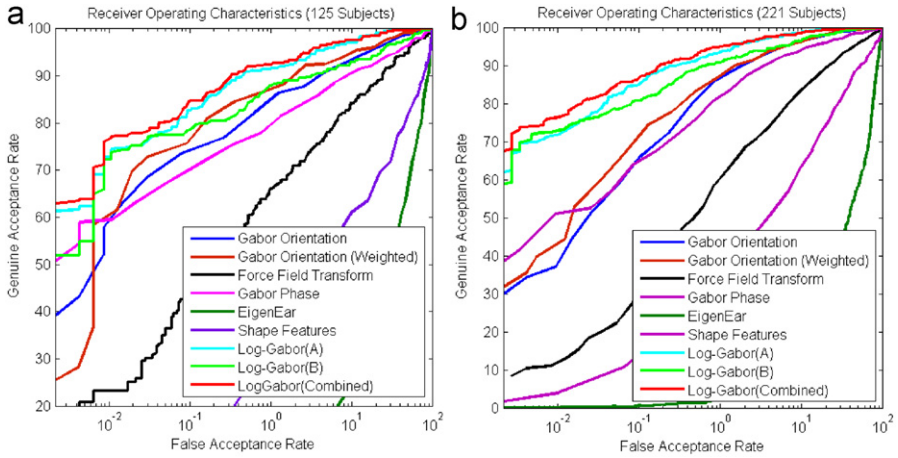

Fig. 18. The receiver operating characteristics for the average authentication performance on the dataset with (a) 125 subjects and (b) 221 subjects.

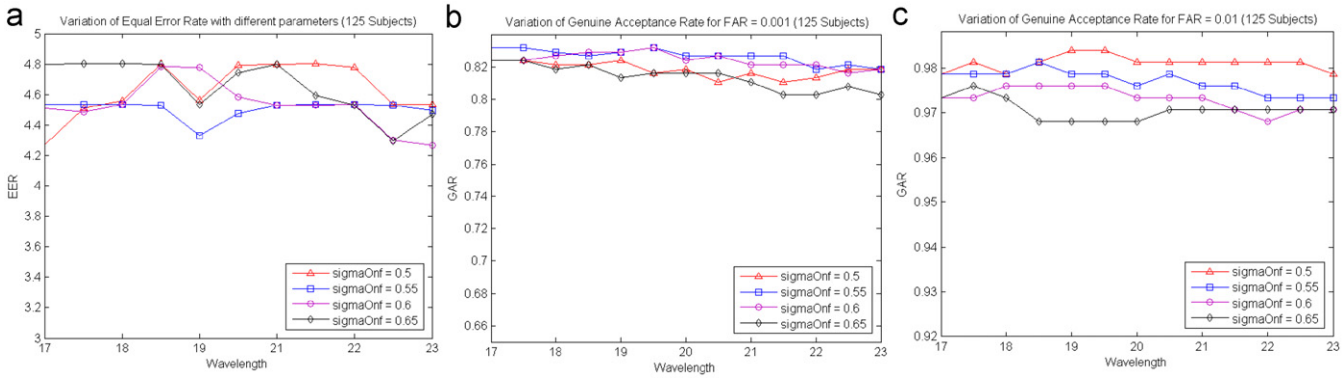

Fig. 19. Sensitivity of the log-Gabor filter parameters for the performance using equal error rate in (a) and genuine acceptance rate in (b) and (c).

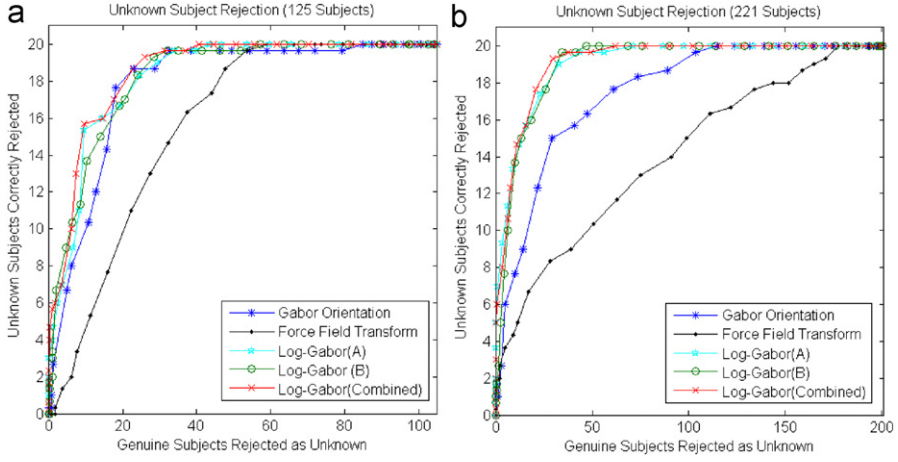

Fig. 20. Performance for the unknown subject rejection using ear images.

segmentation developed in this paper has been highly successful. The manual examination of the localized ear shape boundaries from the corresponding ear images suggested that 400 images (out of total 465 images from 125 subjects, $>85\%$) were satisfactorily/correctly[1] localized. The ear images acquired in this work were acquired in indoor environment where the external illumination variations were limited (reasonable assumption for the civilian deployment of automated ear identification system). However, further efforts are required to develop more effective automated ear segmentation of ear shape region for imaging in outdoor environment.

In this paper we have examined three new approaches for the feature extraction using local gray level phase, local gray level orientation using Gabor filters, and local phase encoding using a pair of log-Gabor filters. The experimental results presented in this paper have illustrated the superiority of log-Gabor based feature extraction and matching approach. Table 2 provides a

[1] The extracted ear shape boundary correctly falls on the curved outer ear shape region, in the clockwise region between the reference points (as shown in Fig. 6).

Please cite this article as: A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.06.005

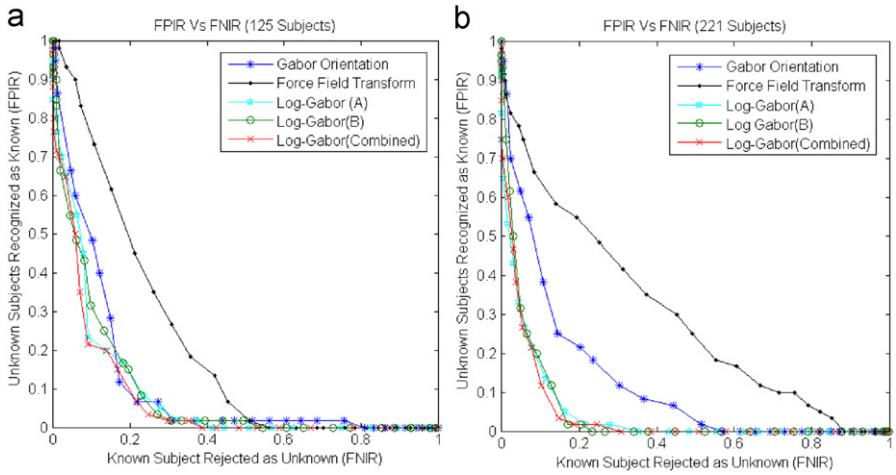

Fig. 21. The FPIR vs. FNIR characteristics from the (a) 125 subject and (b) 221 subject ear database.

Table 2
Summary of related work on personal identification using 2D ear images.

Reference	Ear segmentation	Features	Classifier	Recognition	Verification	Probe database (subjects)	Performance (%)[a]
Mu et al. [15]	Manual	Shape features (outer ear and inner ear)	BP network	Yes	No	77	85 (rank-1)
Hurley et al. [5]	Automated	Force field transform, PCA	*K-NN*	Yes	No	63	99.2 (rank-1)
Abate et al. [21]	Manual[b]	Generic Fourier descriptor	*K-NN*	Yes	No	70	88 (rank-1, 15/30°) 96 (rank-1, 0°)
Chang et al. [7]	Manual	PCA	*K-NN*	Yes	No	111 (probe)	71.6 (rank-1)
Moreno et al. [23]	Manual	Ear profile representing shape and wrinkles	Neural Network	Yes	No	28 (probe)	93 (rank-1)
Nanni et al. [30]	Manual	Color space selection using SFFS	*K-NN*	Yes	Yes	64 (probe)	84 (rank-1)
Abdel Mottaleb et al. [25]	Automated	Hausdorff distance for surface curvature points	*K-NN*	Yes	No	29 (probe)	87.93 (rank-1)
Bustard and Nixon [26]	Automated	Homography transform from SIFT features	*A-NN*	Yes	No	63 (probe)	96 (rank-1)
This paper	Automated	Orthogonal log-Gabor filter Pair	*K-NN*	Yes	Yes	125, 221 (probe)	96.27, 95.93 (rank-1)

[a] Representative best results only (more in the reference).
[b] Also automated ear detection using Haar based object detector.

comparative survey of the related prior work on 2D ear identification reported in the literature. This table also shows the performance achieved relative to the segmentation approach (automated or manual) and the size of probe dataset/subjects employed in the related experiments. It may be observed that the performance of the proposed approach (on a relatively larger database) is superior than related approaches reported in the literature. However, it should be also noted that such one-to-one comparison of these approaches cannot be made as the achieved performances are obtained on different databases, with non-standard experimental configurations/protocols.

6. Conclusions

We have presented a fully automated approach for the personal identification using 2D ear imaging. The proposed approach for the automated ear segmentation using morphological operators and Fourier descriptors has been quite effective in the robust segmentation of the curved region of interest. Another key effort in this paper has been to investigate new feature extraction approaches for the 2D ear images. We have exploited the local orientation features using even Gabor filters and achieved superior performance with prior approaches (using eigen ear, force field transform, shape features). However the best performance, i.e., rank-one recognition accuracy of 96.27% and 95.93%, respectively, on the database of 125 and 221 subjects, was observed from the feature extraction approach using a pair of log-Gabor filters as detailed in this paper. Our experimental results from the authentication experiments and the FPIR vs. FNIR performance have also suggested that the log-Gabor based feature extraction approach outperforms other feature extraction approach considered in this work. One of the key observations of our work is related to the effectiveness of shape features in a completely automated system. Despite our best efforts to achieve the robust localization of curved ear shape and its representation, the achieved performance was poor. Our observations have suggested that the discriminibility of such shape features is quite limited, especially in the presence of large number of subjects as in our work. The ear database acquired in this work from 125 subjects (also 221 subjects), along with our segmented images, is now made publicly available [33] to the researchers. This will help to foster further efforts in the automated ear identification and researchers can reproduce our results to suggest possibly more effective alternatives.

Our efforts to achieve automated ear identification have achieved significant results in this paper. However, further efforts are required to ascertain the effect of pose, occlusion and eyeglasses on the performance. The experimental results presented in this paper have been examined limited orientation and scale

Please cite this article as: A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.06.005

variations present in the employed dataset. Further work is therefore also required to ascertain the performance in presence of large orientation and scale variations in the ear images. Further work should also be directed to improve the automated ear segmentation capability in the indoor environment, and develop ear recognition capability at a distance, in the outdoor environment, where significant changes in the illumination and shadows could pose severe challenges in the robust ear segmentation.

Acknowledgement

This work was supported by the internal competitive research grant from The Hong Kong Polytechnic University (2009–2010), under Grant 4-Z0F3.

References

[1] A. Iannerelli, Ear Identification, Forensic Identification Series, Paramount Publishing Company, Fremount, California, 1989.
[2] M. Burge, W. Burger, "Ear Biometrics," Biometrics: personal identification in networked society, in: A. K. Jain, R.Bolle, S. Pankanti (Eds.), 1998, pp. 273–286.
[3] M. Burge, W. Burger, Ear Biometrics in Machine Vision, in: Proceedings of the 21st Workshop of the Australian Association for Pattern Recognition, 1997.
[4] D.J. Hurley, M.S. Nixon, J.N. Carter, Force field energy functionals for image feature extraction, Image and Vision Computing 20 (5–6) (2002) 311–318.
[5] D.J. Hurley, M.S. Nixon, J.N. Carter, Force field energy functionals for ear biometrics, Computer Vision and Image Understanding 98 (3) (2005) 491–512.
[7] K. Chang, B. Victor, K.W. Bowyer, S. Sarkar, Comparison and combination of ear and face images in appearance-based biometrics, IEEE Transactions on Pattern Analysis Machine Intelligence 25 (8) (2003) 1160–1165.
[10] B. Bhanu, H. Chen, Human ear recognition in 3D, in: Proceedings of the Multimodal User Authentication workshop (MMUA), Santa Barbara, CA, 2003, pp. 91–98.
[11] H. Chen, B. Bhanu, Human ear recognition in 3D, IEEE Transactions on Pattern Analysis Machine Intelligence 29 (4) (2007) 718–737.
[12] D. Fields, Relations between the statistics of natural image and the response properties of cortical cells, Journal of the Optical Society of America 4 (12) (1987) 2379–2394.
[13] J. Daugman, The importance of being random: statistical principles of iris recognition, Pattern Recognition 36 (2003) 279–291.
[14] Z.-X. Xie, Z.-C. Mu, Improved locally linear embedding and its application on multi-pose ear recognition, in: Proceedings of the International Conference on the Wavelet Analysis Pattern Recognition, Beijing, PR China, Nov. 2007, pp. 1367–1371.
[15] Z. Mu, L. Yuan, Z. Xu, D. Xi, S. Qi, Shape and structural feature based ear recognition, in: Sinobiometrics 2004, LNCS, vol. 3338, 2004, pp. 663–670.
[18] P. Yan, K.W. Bowyer, Biometric recognition using 3D ear shape, IEEE Transactions on Pattern Analysis Machine Intelligence 29 (8) (2007) 1297–1308.
[19] B. Arbab-Zavar, M.S. Nixon, On shape-mediated enrolment in ear biometrics, in: Proceedings of the International Symposium on Visual Computing, ISVC'07, Nevada, USA, Nov. 2007.
[20] M. Choras, Ear biometrics based on geometrical method of feature extraction, in: Proceedings of the AMDO 2004, LNCS, vol. 3179, 2004, pp. 51–61.
[21] A.F. Abate, M. Nappi, D. Riccio, S. Ricciardi, Ear recognition by means of a rotation invariant descriptor" in: Proceedings of the ICPR 2006, Hong Kong, 2006.
[22] Carreria-Perpinan, Compression Neural Networks for Feature Extraction: Application to Human Recognition from Ear Images, M.Sc. Thesis, Faculty of Informetics, Technical University of Madrid, Spain 1995.
[23] B. Moreno, A. Aanchez, J.F. Velez, Use of outer ear images for personal identification, in: Proceedings of the 33rd Annual International Carnahan Conference, Madrid, 1999, pp. 469–476.
[25] M. Adbel-Mottaleb, J. Zhou, Human ear recognition from face profile images, in: Proceedings of the ICB 2006, LNCS, vol. 3832, 2006, pp. 786–792.
[26] J.D. Bustard, M.S. Nixon, Towards unconstrained ear recognition from two dimensional images, IEEE Transactions on Systems, Man, and Cybernetics—Part C 40 (3) (2010) 486–494.
[27] A.K. Jain, Y. Chen, M. Demirkus, Pores and ridges: high resolution fingerprint matching using level 3 features, IEEE Transactions on Pattern Analysis Machine Intelligence 29 (2007) 15–27.
[28] W.K. Kong, D. Zhang, Competitive coding scheme for palmprint verification, in: Proceedings of the ICPR, Washington DC, 2004, pp. 1051–xxx.
[29] A. Kumar, Y. Zhou, Human identification using knucklecodes, in: Proceedings of the IEEE Third International Conference on Biometrics: Theory, Applications, and Systems, BTAS '09, Washington D.C., 2009, pp. 1–6.
[30] L. Nanni, A. Lumini, Fusion of color spaces for ear authentication, Pattern Recognition 42 (2009) 1906–1913.
[31] D. Woodard, T. Faltemier, O. Yan, P. Flynn, K. Bowyer, A comparison of 3D biometric modalities, in: Proceedings of the CVPR, CVPR'W06, 2006, pp. 57–61.
[32] T. Theoharis, G. Passalis, G. Toderici, I. Kakadiaris, Unified 3d face and ear recognition using wavelet on geometry images, Pattern Recognition 41 (2008) 796–804.
[33] IIT Delhi Ear Database Version 1, 〈http://webold.iitd.ac.in/~biometrics/Database_Ear.htm〉.
[34] J. Daugman, How iris recognition works, IEEE Transactions on CSVT I (2004) 21–30.
[36] E. de Ves, X. Benavent, G. Ayala, J. Domingo, Selecting the structuring element for morphological texture classification, Pattern Analysis & Applications 9 (1) (2007) 48–51.
[37] A. Kumar, A. Passi, Comparison and combination of iris matchers for reliable personal authentication, Pattern Recognition 43 (3) (2010) 1016–1026.
[38] R.C. Gonzalez, R.E. Woods, Digital Image Processing, 3rd Ed., Pearson, 2008.

About the Author—Ajay Kumar received his Ph.D. degree from the University of Hong Kong in May 2001. From October 2001 to December 2002 he was a Postdoctoral Researcher in the Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong. From 2003 to 2005 he was a Postdoctoral Fellow at the Hong Kong Polytechnic University, where he was in the Department of Computing from April 2004 to January 2005. From 2005 to 2007, he was an Assistant Professor, Department of Electrical Engineering, Indian Institute of Technology Delhi, India. He is currently an Assistant Professor with the Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong. He has been an area-editor for the Encyclopaedia of Biometrics (Springer, 2009) and served on the program committees of several international conferences and workshops in the field of his research interest. He was the program chair of The Third International Conference on Ethics and Policy of Biometrics and International Data Sharing in 2010. His current research interests are on biometrics with the emphasis on hand biometrics, vascular biometrics, iris and multimodal biometrics.

About the Author—Chenye Wu received his Bachelors in Electronics Engineering from Tsinghua University in 2009. He has been working in the department of computing, The Hong Kong Polytechnic University, during his summer internship for three months. His research interests are in the algorithms, theoretical and computational computer science.

Please cite this article as: A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition (2011), doi:10.1016/j.patcog.2011.06.005