Robust Ear Recognition using Gradient Ordinal Relationship Pattern *

Aditya Nigam[1] and Phalguni Gupta[2]

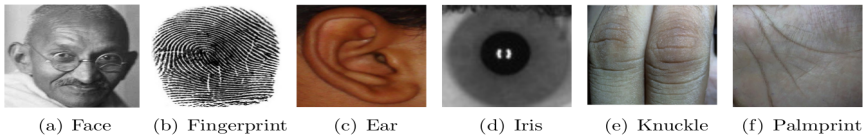
[1] School of Computing and Electrical Engineering,
Indian Institute of Technology Mandi, India - 175005,
aditya@iitmandi.ac.in
[2] Department of Computer Science and Engineering,
Indian Institute of Technology Kanpur, India - 208016
pg@cse.iitk.ac.in

Abstract. A reliable personal recognition based on ear biometrics is highly in demand due to its vast application in automated surveillance, law enforcement *etc.* In this paper a robust ear recognition system is proposed using gradient ordinal relationship pattern. A reference point based normalization is proposed along with a novel ear transformation over normalized ear, to obtain robust ear representations. Ear samples are enhanced using a local enhancement technique. Later a dissimilarity measure is proposed that can be used for matching ear samples. Two publicly available ear databases IITD and UND-E are used for the performance analysis. The proposed system has shown very promising results and significant improvement over the existing state of the art ear systems. The proposed system has shown robustness against small amount of illumination variations and affine transformations due to the virtue of ear transformation and tracking based matching respectively.

1 Introduction

Personal authentication plays an important role in the society. Every application requires at least some level of security to assure personal identity. Hence an automated and accurate human access control mechanism plays an important role in several social applications such as law enforcement, secure banking, immigration control *etc.* Security can be realized at one of the three levels.

1. **Level-1 [Possession]:** The user possesses something which is required to be produced for successful authentication. For example, key of a car or room.
2. **Level-2 [Knowledge]:** The user knows something which is required to be entered correctly for successful authentication. For example, PIN (personal identification number), credit card CVV (card verification value) *etc.*
3. **Level-3 [Biometrics]:** The user owns certain physiological and behavioral characteristics which are required to be acquired and matched for successful authentication. For example, face, iris, fingerprint, signature, gait *etc.*



(a) Face (b) Fingerprint (c) Ear (d) Iris (e) Knuckle (f) Palmprint

(a) Physiological

(a) Signature (b) Gait (c) Voice

(b) Behavioral

Fig. 1. Biometric Traits

Biometrics can be used for personal authentication using physiological (such as face [1, 2], fingerprint [3, 4], iris [5–7], palmprint [8], ear [9, 10], knuckleprint [11–13] *etc.*) and behavioral (such as gait, speech *etc.*) characteristics as shown in Fig. 1, which are assumed to be unique for each individual. However, each trait has its own challenges and trait specific issues hence none of the biometric trait can be considered as the best one. The biometric trait selection completely depends upon the application where ultimately it is going to be deployed. Ear can be considered as a significant biometric because of its passive nature, as it can be acquired without much user cooperation unlike iris, fingerprint. In [14], manually measured twelve features extracted from 10, 000 ear samples suggested that ear may contain unique characteristics. Also it is inherently different from the invasive biometrics such as face which changes with expression, age, illumination, pose and artifacts such as beard, sunglasses *etc.* Moreover the structure and size of a ear is assumed to be invariant for wide age ranges. Ear size is bigger than fingerprint and iris and smaller than face hence can be easily acquired and processed. Still, scale and affine transformation remains to be the major challenges for any ear based recognition system.

In this paper an automated reference point based ear normalization is proposed to handle scale and affine variations. Also an image transformation along with a feature extraction and matching algorithm is proposed that can compliment each other for better performance. The system is tested over two publicly available ear databases *viz.* IITD [15] and UND-E [16]. It has shown significantly better performance than the state of the art ear authentication systems, presented in all these recent ear identification journal works [10], [17] and [15]. The paper is organized as follows. In Section 2 the literature available on ear recognition is reviewed. The proposed ear recognition system is discussed in Section 3. The experimental results are shown in Section 4. Conclusions are discussed in the last Section.

⋆ Authors would like to acknowledge the funding and support provided by IIT, Mandi.

Table 1. Summarized Literature Review

Approach	Number of Images	Performance (Rank 1)
Force Field Transformation and NKFDA [18]	711	75.3 %
Sparse Representation [19]	192	100 %
Locally Linear Embedding (LLE) [20]	1501	80 %
SIFT Landmark Points over Ear Model Face [21]	1060	95.32 %
SIFT features from various color models [22]	800	96.93 %
Partitioned Iterated Function System(PIFS) [23]	228	61 %
Gabor Wavelet and GDA [24]	308	99.1 %
Local Binary Pattern and Wavelet [25]	308	100 %

2 Literature Review

Several ear recognition and authentication systems have been proposed in last decade. Force field transformation [18] is used along with null space based kernel fisher discriminant analysis $NKFDA$ to perform multi-pose ear recognition. Sparse representations [19] are used to develop a dictionary during training stage. A linear system of equations is obtained from probe images and is solved using sparsity of probe image vector and l_1 minimization. Ear recognition is done using local linear embedding LLE and its improved version in [20]. Feature level fusion of multiple ear samples is done for better results in [21]. $SIFT$ features are fused to obtain a single fused template from all training images. In [22] regions with consistent color are used for matching using $SIFT$ features with GMM based skin modeling. Human ear recognition against occlusion ($HERO$) is proposed in [23] which is based on fractals. It can deal with synthetic and natural occlusion. They indexed images using partitioned iterated function system ($PIFS$). Gabor filter [24] based feature extraction and general discriminant analysis GDA are used together for ear recognition.

Some of the current state of the art systems are discussed as follows. In [9], connected component analysis of the graph constructed using edge-map is used to segment ear and SURF features are matched using nearest neighborhood ratio matching. Automated ear segmentation [10] is done using Fourier descriptors and ear identification is performed using log-gabor filters. Later in [17], they have utilized the phase information using 2D quadrature filtering to improve their results. Finally in [15], they have proposed sparse feature coding scheme using localized radon transformation in order to produce their best result. Table 1 summarizes the most significant work done in the field of ear recognition.

3 Proposed Method

The proposed ear recognition system consist of following steps, *viz*. ROI extraction, Reference point detection (Section 3.1), Ear Normalization (Section

(a) Original (b) Thresholded (c) Canny (d) Ear Helix (e) Ref. Points

Fig. 2. Reference Points

3.2), Ear Enhancement (Section 3.3), Ear Transformation (Section 3.4), Feature Extraction (Section 3.5) and Matching (Section 3.6). The ear ROI from any profile face image is cropped using the connected component graph based algorithm proposed in [26]. The cropped ear is automatically normalized using the proposed method that first detect two reference points and later use them for ear normalization. The top and bottom-most ear points are chosen as reference points as shown in Fig. 2(e). The normalized ear is further enhanced and transformed to obtain robust ear representations. Finally the corner features are extracted and matched using the tracking based dissimilarity measure named as Incorrectly Tracked Corners (ITC).

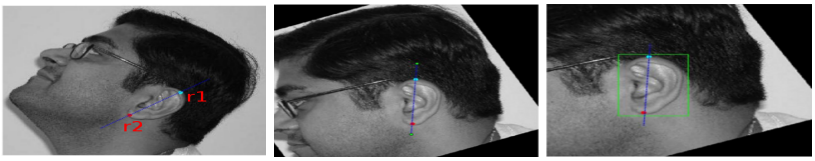
3.1 Reference Point Detection

The major problem in ear recognition is of scale variations. Hence to handle it ear samples are normalized using two reference points r_1 and r_2. Hence these two reference points that are considered as top and bottom ear points (as shown in Fig. 2(e)) are detected automatically. The cropped ear ROI image is preprocessed using an adaptive thresholding [27] and the noise is removed by applying median filtering. The edges are detected using Canny edge detection algorithm [28], and the output is divided into two halves. From both halves the largest contour is extracted, which is considered as the outer helix part of the ear as shown in Fig. 2(d). The two reference points r_1 (over upper helix) and r_2 (over lower helix) are the points which are at maximum distance over these helical structures as shown in Fig. 2(e) and Eq. 1. In Eq. 1, x, y represents all points belonging to upper and lower helix respectively, r_1 and r_2 are the required reference points and $||x, y||$ represents the euclidean distance between x and y point.

$$\{r_1, r_2\} = argMax_{\forall (x \in UP.Helix) \ and \ (y \in LW.Helix)} ||x, y|| \tag{1}$$

3.2 Ear Normalization

The issues of affine transformations such as rotation, scaling and alignment, are overpowered by the using the proposed automated ear normalization. All database profile face images are scaled to a predefined size and are registered using two reference points *viz.* (r_1) and (r_2), which are detected as discussed



(a) Original (I_{Orig}) (b) Nor. Face (I_t) (c) Ear ROI (I_e)

Fig. 3. Ear Normalization

above and shown in Figs. 2(e), 3(a). The reference points are detected in cropped ear ROI but actual profile image is normalized to ensure that:

1. The points r_1 and r_2 should lie on same coordinates in all the images (*i.e.* (x_1^t, y_1^t) and (x_2^t, y_2^t) respectively).
2. The line joining r_1 and r_2 should be vertically aligned as shown in Fig. 3(c).

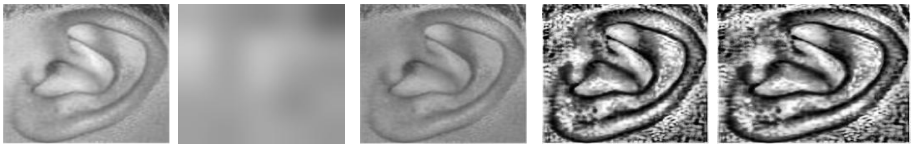
Let the co-ordinates of r_1 and r_2 are (x_1^o, y_1^o) and (x_2^o, y_2^o) respectively, in original image I_{orig} as shown in Fig. 3(a). The image I_{orig} is scaled to I_s by a scaling factor $S = \frac{d}{distance(r_1, r_2)}$ with respect to r_1, so as to set a predefined fixed distance (d) between the points r_1 and r_2 using the scaling matrix $[T_s]$, as given in Eq. (2). The scaled image I_s is then rotated to I_r by an angle ϕ with respect to r_1, where ϕ is the angle between the vertical direction (*i.e.* y-axis) and the line-segment r_1 and r_2 using the rotation matrix $[T_r]$, as given in Eq. (2). Now I_r is translated to I_t (*i.e* Nor. Face, Fig. 3(b)), by $t_x = x_1^t - x_1^o$ and $t_y = y_1^t - y_1^o$ units, in x and y directions using the translation matrix $[T_t]$. The combined image transformation matrix $[T]$ is given below.

$$T = \underbrace{\begin{bmatrix} S & 0 & x_1^o \cdot (1-S) \\ 0 & S & y_1^o \cdot (1-S) \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Scaling Matrix } [T_s] \; w.r.t. \text{ point } r_1} \underbrace{\begin{bmatrix} cos\phi & -sin\phi & x_1^o \cdot (1-cos\phi) + y_1^o \cdot sin\phi \\ sin\phi & cos\phi & y_1^o \cdot (1-cos\phi) - x_1^o \cdot sin\phi \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Rotation Matrix } [T_r] \; w.r.t. \text{ point } r_1} \underbrace{\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Translation } [T_t]} \tag{2}$$

In order to extract a consistent ear ROI, height-width ratio (δ=1.4) is used that is computed heuristically from a set of randomly selected 100 images (50 images each from IITD and UND-E dataset). As the height of a ear is normalized to (d in I_t), by using the above derived transformation matrix T (as given in Eq. (2)) over I_{orig}, the width of the ear box to be cropped can be computed as $w = \frac{\delta}{d} + offset$. Here $offset$ is the left margin along each box side with the line joining the reference points exactly at the midpoint of the box, as shown in Fig.3(c).

3.3 Ear Enhancement

The extracted region of interest (ROI) of an ear contains texture information but generally is of poor contrast. Suitable image enhancement technique is required

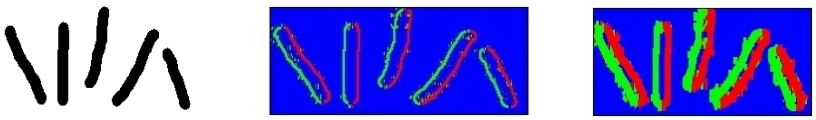


(a) Original (b) Bg Illum. (c) Uni. Illum. (d) Enhanced (e) Noise Removal

Fig. 4. Ear Image Enhancement

to apply on the ROI. The ear texture is enhanced in such a way that it increases its richness as well as its discriminative power. The ear ROI is divided into blocks and the mean of each block is considered as the coarse illumination of that block. This mean is expanded to the original block size as shown in Fig. 4(b). Selection of block size plays an important role. It should be such that the mean of the block truly represents the illumination effect of the block. So, larger block may produce improper estimate. Since the ear sample images are of size 160×200, it is observed that a block size of 40×40 is the best choice for our experiment. The estimated illumination of each block is subtracted from the corresponding block of the original image to obtain the uniformly illuminated ROI as shown in Fig. 4(c). The contrast of the resultant ROI is enhanced using Contrast Limited Adaptive Histogram Equalization ($CLAHE$) [29]. It removes the artificially induced blocking effect using bilinear interpolation and enhances the contrast of image without introducing much external noise. Finally, Wiener filter [30] is applied to reduce constant power additive noise to obtain the enhanced ear texture as shown in Fig. 4(e).

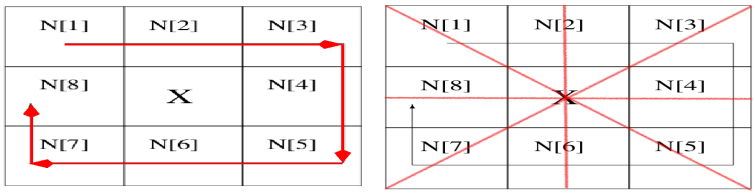
3.4 Ear Transformation



(a) Image (b) Kernel=3×3 (c) Kernel=9×9

Fig. 5. Red: -ve grad; Green: +ve grad.; Blue: zero grad (Kernel size represents the size of sobel kernel used to compute gradient.)

The normalized and enhanced ear samples are transformed using two proposed encoding schemes Gradient Ordinal Relation Pattern ($GORP$) and STAR GORP ($SGORP$). The gradient of any edge pixel is positive if it lies on an edge created due to light to dark shade (*i.e. high to low gray value*) transition else it will be having negative gradient or zero value. Hence all the edge pixels can be divided into three classes of $+ve$, $-ve$ or zero gradient values as shown in



(a) *GORP* (b) *SGORP*

Fig. 6. Neighborhood structure used to compute the gradient and the robust encodings.

Fig. 5(b,c). The *sobel* x-direction kernel of size 3×3 and 9×9 are applied to obtain Fig. 5(b,c) respectively. Bigger size kernel produces coarse level features. The *sobel* kernel lacks rotational symmetry hence more consistent *scharr* kernel [31] which is obtained by minimizing angular error is applied. This gradient augmented information of an edge pixel can be more discriminative and robust. The proposed encoding schemes *GORP* and *SGORP* precisely uses this information to calculate a 8-bit code for each pixel by using x and y-direction derivatives of its 8 neighboring pixels to obtain four codes *viz.* $vcode^{GORP}$, $hcode^{GORP}$, $vcode^{SGORP}$ and $hcode^{SGORP}$ as shown in Figs. 7(b,c,d,e) respectively. The vertical/horizontal code (*i.e.* *vcode* and *hcode*) signifies that whether x-direction or y-direction derivatives are used while encoding.

Let $P_{i,j}$ be the $(i,j)^{th}$ pixel of an ear sample P and $N[l]$, $l = 1, 2, ...8$ (as shown in Fig. 6(a)) are the gradients of 8 neighboring pixels centered at pixel $P_{i,j}$ that are obtained by applying *scharr* kernel.

[a] GORP based Encoding: An eight bit encoding *viz.* *gorp_code* for every pixel is computed whose k^{th} bit can be defined as :

$$gorp\_code[k] = \begin{cases} 1 & \textbf{if } N[k] > 0 \\ 0 & \textbf{otherwise} \end{cases} \tag{3}$$

In $vcode^{GORP}$ or $hcode^{GORP}$ as shown in Fig. 7 (b,c) every pixel is represented by its *gorp_code* instead of its gray value which is obtained using the gradient of its 8 neighboring pixels computed using x and y direction *scharr* kernel respectively.

(a) Original (b) $vcode^{GORP}$ (c) $hcode^{GORP}$ (d) $vcode^{SGORP}$ (e) $hcode^{SGORP}$

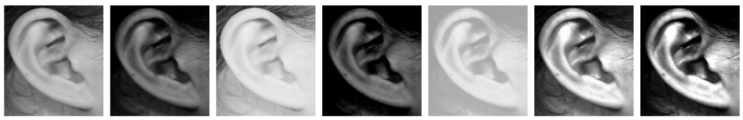
Fig. 7. Ear Transformation Using Local Gradient

[b] SGORP based Encoding: The *sgorp_code* encodes the ordinal relationships between diagonally opposite neighbors along with top and bottom one (as shown in Fig. 6(b) in red color). Due to this star structure it is termed as STAR GORP (*i.e SGORP*). In $vcode^{SGORP}$ or $hcode^{SGORP}$ as shown in Fig.7 (d,e) every pixel is represented by its *sgorp_code* instead of its gray value. It is a 4 bit code represented in 8 bits as follows:

$$\begin{aligned} sgorp\_code[1] = sgorp\_code[2] = 1 \ \ \textbf{if} \ \ abs(N[1] - N[5]) > T \ \textbf{else} \ 0 \\ sgorp\_code[3] = sgorp\_code[4] = 1 \ \ \textbf{if} \ \ abs(N[2] - N[6]) > T \ \textbf{else} \ 0 \\ sgorp\_code[5] = sgorp\_code[6] = 1 \ \ \textbf{if} \ \ abs(N[3] - N[7]) > T \ \textbf{else} \ 0 \\ sgorp\_code[7] = sgorp\_code[8] = 1 \ \ \textbf{if} \ \ abs(N[4] - N[8]) > T \ \textbf{else} \ 0 \end{aligned} \tag{4}$$

where T, is empirically determined threshold that depends upon the size of *scharr* kernel used to compute gradient. The pattern of edges within a neighborhood can be assumed to be robust; hence each pixel's encodings (*i.e. gorp_code* or *sgorp_code*) are considered, that only used sign of the derivative within its specified neighborhood which is their ordinal relationships. Hence they ensures the robustness of the proposed transformation under illumination variation.

Justifications: The subsequent samples of same ear will be having varying illumination and affine transformations. The proposed $GORP$ and $SGORP$ transformation can handle illumination variations. In Fig. 8(a) the first ear sample is shown in six significantly different illumination conditions that are artificially created to observe the robustness of the proposed transformation. The Figs. 8(b) and 8(c) shows their corresponding $hcode^{GORP}$ and $vcode^{GORP}$. One can clearly observe that they are not varying much even under such drastic illumination variations.



(a) An ear image in seven different illumination conditions

(b) Their corresponding $hcode^{GORP}$

(c) Their corresponding $vcode^{GORP}$

Fig. 8. Robustness of $GORP$ against varying Illumination

3.5 Feature Extraction

We have considered corner points as features because of their repeatability and discrimination. The autocorrelation matrix M [32], can be used to calculate corner points that are having strong orthogonal derivatives. The matrix M is a 2×2 matrix and hence it can have two eigen values λ_1 and λ_2 such that $\lambda_1 \geq \lambda_2$ with e_1 and e_2 as the corresponding eigenvectors. All pixels having $\lambda_2 \geq T_q$ (smaller eigen value greater than a threshold) are considered as corner points.

3.6 Feature Matching

Tracking algorithms can be used to achieve robustness against affine transformations as it assumes that any feature can move within a small neighborhood. Tracking can also handle partial ears, as they are occluded mostly by hairs or ear-rings. They can easily be detected and masked using skin-tone detection and feature over them are ignored while tracking. A constrained version of KL tracking [33], has been used to perform matching between two ear samples E_a and E_b. The KL tracking make use of three assumption namely brightness consistency, temporal persistence and spatial coherency as defined below, hence its performance depends completely on how well these three assumptions are satisfied.

Let there be a feature at location (x, y) at a time instant t with intensity $I(x, y, t)$ and this feature has moved to the location $(x + \delta x, y + \delta y)$ at the time instant $t + \delta t$. Three basic assumptions that are used by KL Tracking to perform tracking successfully are :

- **Brightness Consistency**: Features in a frame do not change much for small change in the value of δt, *i.e*

$$I(x, y, t) \approx I(x + \delta x, y + \delta y, t + \delta t) \tag{5}$$

- **Temporal Persistence**: Features in a frame moves only within a small neighborhood. Using the Taylor series and neglecting the high order terms, one can estimate $I(x + \delta x, y + \delta y, t + \delta t)$ with the help of Eq. (5) as :

$$I_x V_x + I_y V_y = -I_t \tag{6}$$

  where V_x, V_y are the components of the optical flow velocity for pixel $I(x, y, t)$ and I_x, I_y and I_t are the derivatives in the corresponding directions.
- **Spatial Coherency**: Estimating unique flow vector from Eq. 6 for every feature point is an ill-posed problem. Hence KL tracking estimates the motion of any feature by assuming local constant flow (i.e a patch of pixels moves coherently). An over-determined system of linear equations have been finally obtained and solved using least square method to estimate the flow vector for each pixel. Finally this vector is used to estimate the new position of that feature.

Justification for Tracking based Matching: In this work we expect that these assumptions are more likely to be satisfied while tracking is performed between features of same subject (genuine matching) and degrades substantially for others (imposter matching). The above fact hold because brightness consistency is assured by $GORP$ and $SGORP$ transformation while the temporal persistence and spatial coherency are satisfied only for genuine matching. Therefore one can infer that the tracking performance of KL tracking algorithm will be good (or at-least better) for genuine matching as compared to imposter.

However, all the tracked corner features may not be the true matches because of the noise, local non-rigid distortions and less difference in inter class matching as compared to intra class matching. Hence a notion of consistent optical flow is proposed to handle this.

Consistent Optical Flow: The true corner matches have their optical flow that can be aligned with the actual affine transformation between the images being matched. Hence the estimated optical flow direction is quantized into eight directions (*i.e.* at an angular difference of $\frac{\pi}{8}$) and the most consistent direction (MCD) is selected as the one which has the maximum number of correctly tracked corner features. Any corner matching pair (*i.e* corner and its corresponding corner) having optical flow direction other than the most consistent direction (MCD) is considered as false matching pair and have to be discarded.

A dissimilarity measure ITC (Incorrectly Tracked Corners) has been proposed to estimate the KL-tracking performance by evaluating simple geometrical and statistical quantities defined as:

Locality: Euclidean distance between any corner and its estimated tracked location should be less than or equal to an empirically selected threshold TH_d. The parameter TH_d depends upon the amount of translation and rotation in the sample images. High TH_d signifies more translation.

Dissimilarity: The tracking error is defined as pixel-wise sum of absolute difference between a local patch centered at current corner feature and that of its estimated tracked location patch, that should have to be less than or equal to TH_e. This error should have to be less than or equal to an empirically selected threshold TH_e. The parameter TH_e ensures that the matching corners must have similar neighborhood patch around it.

Correlation: Phase only correlation [34] between a local patch centered at any feature and that of its estimated tracked location patch should be at-least equal to an empirically selected threshold TH_{cb}.

Any corner is considered as tracked successfully if it satisfies the above defined three constraints. These three parameters (*viz.* TH_d, TH_e and TH_{cb}) are experimentally determined over a small validation set (including only 200 images, 100 each from both dataset) and are optimized $w.r.t$ performance. The values for which the optimized performance is achieved are $TH_d = 13$, $TH_e = 750$ for both $GORP$ and $SGORP$ databases while $TH_{cb} = 0.1$ for $GORP$ and 0.4 for $SGORP$. The $SGORP$ transformation requires more correlation because it is more sparse as compared to $GORP$ as shown in Fig. 7 here correlation and

Algorithm 1 $ITC(E_a, E_b)$

Require:

(a) Let E_a and E_b be two ear images and ($I_A^{v^L}$, $I_B^{v^L}$), ($I_A^{h^L}$, $I_B^{h^L}$), ($I_A^{v^{SL}}$, $I_B^{v^{SL}}$) and ($I_A^{h^{SL}}$, $I_B^{h^{SL}}$) are their four corresponding $vcode^{GORP}$, $hcode^{GORP}$, $vcode^{SGORP}$ and $hcode^{SGORP}$ pairs respectively.

(b) $N_A^{v^L}, N_A^{h^L}, N_A^{v^{SL}}$, and $N_A^{h^{SL}}$ are the number of corners in $I_A^{v^L}, I_A^{h^L}, I_A^{v^{SL}}$, and $I_A^{h^{SL}}$ respectively.

Ensure: Return $ITC(E_a, E_b)$.

1: Track individually corners within the 4 ear encoding pairs ($I_A^{v^L}$, $I_B^{v^L}$), ($I_A^{h^L}$, $I_B^{h^L}$), ($I_A^{v^{SL}}$, $I_B^{v^{SL}}$) and ($I_A^{h^{SL}}$, $I_B^{h^{SL}}$) .

2: Obtain four set of corners that are successfully tracked in,
(a) ($I_A^{v^L}$, $I_B^{v^L}$) pair tracking (*i.e.* stc^{v^L}),
(b) ($I_A^{h^L}$, $I_B^{h^L}$) pair tracking (*i.e.* stc^{h^L}),
(c) ($I_A^{v^{SL}}$, $I_B^{v^{SL}}$) pair tracking (*i.e.* $stc^{v^{SL}}$),
(d) ($I_A^{h^{SL}}$, $I_B^{h^{SL}}$) pair tracking (*i.e.* $stc^{h^{SL}}$) ,
that have their tracked position within TH_d and their local patch dissimilarity under TH_e also the patch-wise correlation is at-least equal to TH_{cb}.

3: Quantize optical flow direction for each successfully tracked corner set into only eight directions (*i.e.* at an interval of $\frac{\pi}{8}$) to obtain 4 histograms H^{v^L}, H^{h^L}, $H^{v^{SL}}$ and $H^{h^{SL}}$ using these four corner set $stc^{v^L}, stc^{h^L}, stc^{v^{SL}}$ and $stc^{h^{SL}}$ respectively as computed above.

4: For each histogram, out of 8 bins the bin (*i.e.* direction) having the maximum corners is considered as the consistent optical flow direction. The maximum value obtained from each histogram is termed as corners tracked correctly (*i.e.* *ctc* value) represented as $ctc^{v^L}, ctc^{h^L}, ctc^{v^{SL}}$ and $ctc^{h^{SL}}$.

5: **return** $ITC(E_a, E_b)=1-\frac{\frac{ctc^{v^L}}{N_A^{v^L}}+\frac{ctc^{h^L}}{N_A^{h^L}}+\frac{ctc^{v^{SL}}}{N_A^{v^{SL}}}+\frac{ctc^{h^{SL}}}{N_A^{h^{SL}}}}{4}$;

path wise error are computed using a block size of 5×5. Let E_a and E_b are two ear sample images that have to be matched and ($I_A^{v^L}$, $I_B^{v^L}$), ($I_A^{h^L}$, $I_B^{h^L}$), ($I_A^{v^{SL}}$, $I_B^{v^{SL}}$) and ($I_A^{h^{SL}}$, $I_B^{h^{SL}}$) are their four corresponding $vcode^{GORP}$, $hcode^{GORP}$, $vcode^{SGORP}$ and $hcode^{SGORP}$ pairs respectively. The Algorithm 1 can be used to compute the $ITC(E_a, E_b)$ based dissimilarity score between two ear sample images E_a and E_b.

4 Experimental Analysis

4.1 Database

The proposed system is tested over two widely used publicly available ear databases IITD [10, 17, 15] and UND-E [16]. The IITD ear database have 493 images collected from 125 subjects (age range 14-58 years) over a period of 9 months in indoor environment with out extra illumination. Each subject has given at-least

3 left ear images. The UND-E database consists of 443 profile face images collected from 114 subjects with 3 to 9 images per subject. Images are collected on different days with different pose and illumination. It can be noted that there exist a huge intra-class variation in these images due to pose variation and different imaging conditions.

4.2 Testing Strategy

The testing strategy used in this work is exactly the same as adopted in [10], [17] and [15] so as to compare our result against these state of the art systems. The test protocol A (*i.e.* single image in training and testing) computes the average of 3 tests in which only initial 3 images per subject are used with one as the *test image* while remaining images as *training images* (one by one). The test protocol B (*i.e.* all to all) computes the average performance when all images from all subjects are used as *training images* (except itself). For IITD dataset using protocol A, 125×3 and $124 \times 125 \times 3$ genuine and imposter matchings are considered respectively while using protocol B, 493 genuine and 124×493 imposter matchings are performed. Similarly for UND-E dataset using protocol A, 110×3 and $109 \times 110 \times 3$ genuine and imposter matchings are considered respectively while using protocol B, 443 genuine and 109×443 imposter matchings are performed.

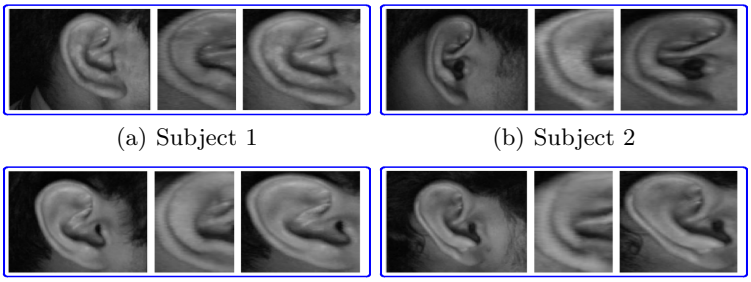
4.3 Performance Analysis

The performance of the system is measured using correct recognition rate (CRR) in case of identification and equal error rate (EER) for verification. The CRR (*i.e.* the **Rank 1** accuracy) of any system is defined as the ratio of the number of correct (Non-False) top best match of ear ROI and the total number of ear ROI in the query set. At any given threshold, the probability of accepting the impostor is known as false acceptance rate (FAR) and the probability of rejecting the genuine user known as false rejection rate (FRR). Equal error rate (EER) is the value of FAR for which FAR and FRR are equal.

$$EER = \{FAR | FAR = FRR\} \tag{7}$$

For all these experimentations average results in terms of EER, CRR and receiver operating characteristics (ROC) [35] curves are reported in Table 2 and Fig. 10. In Table 2 the proposed system is compared with six existing most recent state of the art algorithms proposed in [10], [17] and [15] and is found to be performing much better. The proposed system outperforms non negative formulation (NNG) [15] in both CRR and EER parameters over both databases and using both testing protocol A and B. The prime reasons behind the proposed system's superior performance includes :

– In [10], [17], [15] they have only considered left half of the ear and discarded most of the right half as shown in Fig. 9. But in this work we have utilized the information from both halves of ear ROI.



(a) Subject 1 (b) Subject 2

(c) Subject 3 (d) Subject 4

Fig. 9. Ear Images. Each subject's first image is the raw ear image, second is used by [15] while third image which is full ear ROI is used in this work.

Algorithm	**Protocol A**		**Protocol B**	
	IITD	**UND**	**IITD**	**UND**
	CRR%,EER%	**CRR%,EER%**	**CRR%,EER%**	**CRR%,EER%**
Hessian[15]	95.47, 5.28	86.06, 9.84	95.94, 4.64	85.91, 9.97
LG[10]	95.73, 4.32	87.88, 8.18	95.74, 4.06	88.91, 7.40
Q.Log.G[17]	96.53, 3.73	90.30, 7.27	96.55, 3.78	90.99, 7.16
LRT[15]	94.40, 4.27	84.84, 7.07	94.93, 3.87	84.76, 7.01
s-LRT[15]	97.07, 2.38	91.52, 5.45	97.57, 2.03	92.61, 5.08
NNG[15]	96.80, 1.87	91.52, 5.15	97.16, 1.62	92.38, 5.31
Proposed	**98.93, 1.05**	**98.18, 2.42**	**99.20, 1.43**	**99.20, 4.78**

Table 2. Comparative Performance Analysis

- In this work $GORP$ and $SGORP$ based transformations have been proposed that can handle some amount of illumination variations in the subsequent ear samples as shown in Fig. 8.
- The ear samples are enhanced so as to obtain robust texture information as shown in Fig. 4.
- Matching is done using corner feature tracking, hence it can handle some amount of affine transformations.
- To handle partial and occluded ear by hairs or ear ring, simple Gaussian model based skin tone detection [36] is done and features over occluded regions are discarded while tracking.

From Table 2, one can observe that the CRR as well as EER of the proposed system has been found to be much better than the non negative formulation (NNG) [15] in the case of UND-E database because the proposed systems can handle affine variations better than NNG. Also database enlargement has not been done as performed in [10], [17] and [15] which makes the system very efficient.

In Fig. 10 the ROC curves that plots FAR Vs FRR for all thresholds, are plotted for the proposed ear recognition system. Since FAR and FRR both are

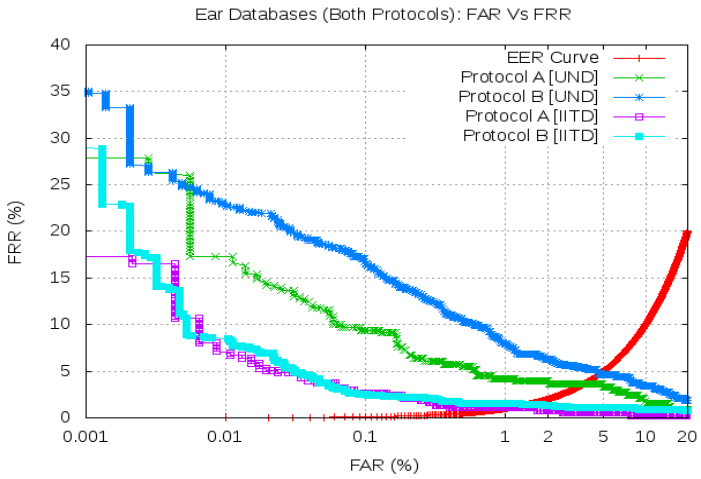


Fig. 10. The Average ROC for IITD and UND Databases. (x-axis is in log scale)

error hence the curve which is lower and towards both axis is considered as better. The curve shown in red color is the EER curve and the point where it intersect any curve is the equal error rate of that system. One can clearly observe that the performance of the propose system is better over IIT ear database under either testing protocol. This is because UND-E ear database is much more challenging and contains ear samples with in-plane rotations.

5 Conclusion

In this paper an end to end automated ear based recognition system is proposed. All steps in the proposed ear recognition system *viz*. ROI extraction, Reference point detection, Ear Normalization, Ear Enhancement, Ear Transformation, Feature Extraction and Matching are proposed and discussed. Two publicly available ear databases IITD [15] and UND-E [16] are considered for performance analysis. The proposed system has shown very promising results and significant improvement over the six recently proposed state of the art ear authentication systems [15], [10] and [17]. The superior performance is achieved by utilizing full (both halves) of ear ROI's and by using the robust encoding scheme ($GORP$ and $SGORP$) along with the tracking based measure that can handle affine transformations effectively. The partial ear and occlusion due to hair and ear ring are handled by using Gaussian model based skin tone detection.

References

1. Nigam, A., Gupta, P.: A new distance measure for face recognition system. In: International Conference on Image and Graphics. ICIG (2009) 696–701

2. Nigam, A., Gupta, P.: Comparing human faces using edge weighted dissimilarity measure. In: International Conference on Control, Automation, Robotics and Vision. ICARCV (2010) 1831–1836
3. Cappelli, R., Ferrara, M., Maltoni, D.: Minutia cylinder-code: A new representation and matching technique for fingerprint recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, **32** (2010) 2128 –2141
4. Singh, N., Nigam, A., Gupta, P., Gupta, P.: Four slap fingerprint segmentation. In: International Conference on Intelligent Computing, ICIC. (2012) 664–671
5. Nigam, A., Gupta, P.: Iris recognition using consistent corner optical flow. In: Asian Conference on Computer Vision, ACCV. (2012) 358–369
6. Nigam, A., Anvesh, T., Gupta, P.: Iris classification based on its quality. In: 9th International Conference on Intelligent Computing, ICIC. (2013) 443–452
7. Bendale, A., Nigam, A., Prakash, S., Gupta, P.: Iris segmentation using improved hough transform. In: International Conference on Intelligent Computing, ICIC. (2012) 408–415
8. Nigam, A., Gupta, P.: Palmprint recognition using geometrical and statistical constraints. In: 2nd International Conference on Soft Computing for Problem Solving, (SocProS). (2012) 1303–1315
9. Prakash, S., Gupta, P.: An efficient ear recognition technique invariant to illumination and pose. Telecommunication Systems **52** (2013) 1435–1448
10. Kumar, A., Wu, C.: Automated human identification using ear imaging. Pattern Recognition **45** (2012) 956–968
11. Nigam, A., Gupta, P.: Finger knuckleprint based recognition system using feature tracking. In: Chinese Conference on Biometrics Recognition, CCBR. (2011) 125–132
12. Badrinath, G.S., Nigam, A., Gupta, P.: An efficient finger-knuckle-print based recognition system fusing sift and surf matching scores. In: International Conference on Information and Communications Security, ICICS. (2011) 374–387
13. Nigam, A., Gupta, P.: Quality assessment of knuckleprint biometric images. In: International Conference on Image Processing, ICIP. (2013) 4205–4209
14. Iannarelli, A.: Ear Identification. Paramount Publishing Company (1989)
15. Kumar, A., Chan, T.S.: Robust ear identification using sparse representation of local texture descriptors. Pattern Recognition **46** (2013) 73–85
16. Chang, K., Bowyer, K.W., Sarkar, S., Victor, B.: Comparison and combination of ear and face images in appearance-based biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 1160–1165
17. Chan, T.S., Kumar, A.: Reliable ear identification using 2-d quadrature filters. Pattern Recognition Letters **33** (2012) 1870–1881
18. Dong, J., Mu, Z.: Multi-pose ear recognition based on force field transformation. In: Second International Symposium on Intelligent Information Technology Application, IITA. Volume 3. (2008) 771–775
19. Naseem, I., Togneri, R., Bennamoun, M.: Sparse representation for ear biometrics. In: Advances in Visual Computing. Volume 5359 of LNCS. (2008) 336–345
20. Xie, Z., Mu, Z.: Ear recognition using lle and idlle algorithm. In: 19th International Conference on Pattern Recognition, ICPR. (2008) 1–4
21. Badrinath, G., Gupta, P.: Feature level fused ear biometric system. In: 7th International Conference on Advances in Pattern Recognition, ICAPR. (2009) 197–200
22. Kisku, D.R., Mehrotra, H., Gupta, P., Sing, J.K.: Sift-based ear recognition by fusion of detected keypoints from color similarity slice regions. In: International Conference on Advances in Computational Tools for Engineering Applications, ACTEA, IEEE (2009) 380–385

23. De Marsico, M., Michele, N., Riccio, D.: Hero: Human ear recognition against occlusions. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW. (2010) 178–183
24. Wang, X., Yuan, W.: Gabor wavelets and general discriminant analysis for ear recognition. In: 8th World Congress on Intelligent Control and Automation, WCICA. (2010) 6305–6308
25. Zhi-qin, W., Xiao-dong, Y.: Multi-scale feature extraction algorithm of ear image. In: International Conference on Electric Information and Control Engineering, ICEICE. (2011) 528–531
26. Prakash, S., Jayaraman, U., Gupta, P.: Connected component based technique for automatic ear detection. In: 16th IEEE International Conference on Image Processing, ICIP. (2009) 2741–2744
27. Shafait, F., Keysers, D., Breuel, T.M.: Efficient implementation of local adaptive thresholding techniques using integral images. In: Electronic Imaging, International Society for Optics and Photonics (2008) 681510–681510
28. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-8** (1986) 679–698
29. Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. Computer Vision, Graphics and Image Processing **39** (1987) 355 – 368
30. Wiener, N.: Extrapolation, Interpolation, and Smoothing of Stationary Time Series. The MIT Press (1964)
31. Scharr, H.: Optimal operators in digital image processing. PhD thesis (2000)
32. Shi, J., Tomasi: Good features to track. In: Computer Vision and Pattern Recognition, CVPR. (1994) 593–600
33. Lucas, B.D., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. In: 7th International Joint Conference on Artificial Intelligence, IJCAI. (1981) 674–679
34. Meraoumia, A., Chitroub, S., Bouridane, A.: Fusion of finger-knuckle-print and palmprint for an efficient multi-biometric system of person recognition. In: IEEE International Conference on Communications, ICC. (2011)
35. Badrinath, G., Gupta, P.: Palmprint based recognition system using phase-difference information. Future Generation Computer Systems **28** (2012) 287 – 305
36. Cai, J., Goshtasby, A.: Detecting human faces in color images. In: Image and Vision Computing 18(1). (1999) 63–75