Human Identification by Ear Images using SIFT Algorithm

Himanshu Maurya[1], Shikha Maurya[2]

[1]National Institute of Technology, Patna, India

[2]A.B.V. Indian Institute of Information Technology and Management, Gwalior, India

Abstract: ***This paper presents a method for extracting distinctive invariant features from ear images that can be used to perform reliable matching between different views of an ear. It shows the extraction of features from an ear image and also the results of it are matching with other samples of both same and different subjects. Each of these feature vectors is supposed to be distinctive and invariant to any scaling, rotation or translation of the image.***

Keywords: Biometric**,** Feature extraction, SIFT, Matching.

1. Introduction

Ear biometric system uses features of the ear (Figure 1) to identify or verify an individual's identity. The comparison is based on variations of the complex structure of the ear. Ear growth after the first four months of birth is proportional to gravity can cause the ear to undergo stretching. Although humans lack the ability to recognize one another from ears, ears have both reliable and robust features which are extractable from a distance. Other biometrics such as fingerprints, hand geometry and retinal scans require close contact and may be considered invasive by users. Study by Iannarelli showed the feasibility of ear biometrics [1]. Possible applications for ear biometrics are ATM machines, evidence surveillance and recognition systems or access to any restricted area.

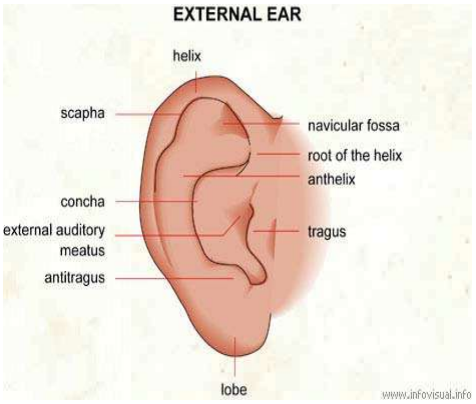

Figure 1: Structure of external ear

2. Literature Survey

There has been quite a bit of work on ear biometric. Iannarelli system is anthropometric technique based upon the 12 ear measurements. It requires exact alignment and normalization of the ear photo and allows comparable measurements. All measurements are relative to an origin which is a point chosen on the image. Hui Chen and Bir Bhanu introduced a two-step ICP (Iterative Closest Point) algorithm for matching 3D ears and Contour Matching for 3D Ear Recognition. Eigen ear and PCA based approach have also been used for ear identification [2].

3. Related Work

David G. Lowe [3] used the SIFT features for object recognition. According to their work, the invariant features extracted from images can be used to perform reliable matching between different views of an object or scene. The features have been shown to be invariant to image rotation and scale and robust across a substantial range of affine distortion, addition of noise, and change in illumination. The approach is efficient on feature extraction and has the ability to identify large numbers of features. Our implementation focuses on deriving SIFT features from an ear image and trying using these features to perform human identification.

4. SIFT (Scale Invariant Feature Transform)

It is a method of finding distinctive and invariant features. The algorithm has four stages.

4.1. Scale Space Extrema Detection

This is the stage where the interest points, which are called key points in the SIFT framework, are detected. For this, the image is convolved with Gaussian filters at different scales, and then the difference of successive Gaussian-blurred images is taken. Key points are then taken as maxima/minima of the Difference of Gaussians (DoG) that occur at multiple scales. Specifically, a DoG image $D(x, y, \sigma)$ is given by,

$$D(x, y, \sigma) = L(x, y, k_i\sigma) - L(x, y, k_j\sigma) \quad (1)$$

Where $L(x, y, k\sigma)$ is the original image $I(x, y)$ convolved with the Gaussian blur $G(x, y, k\sigma)$ at scale kσ, i.e.

$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y) \quad (2)$$

Where,

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \quad (3)$$

Hence a DoG image between scales $k_i\sigma$ and $k_j\sigma$ is just the difference of the Gaussian-blurred images at scales $k_i\sigma$ and $k_j\sigma$. For scale-space extrema detection in the SIFT

algorithm, the image is first convolved with Gaussian-blurs at different scales. The convolved images are grouped by octave, and the value of k_i is selected so that we obtain a fixed number of convolved images per octave. Then the Difference-of-Gaussian images are taken from adjacent Gaussian-blurred images per octave.

Once DoG images have been obtained, key points are identified as local minima/maxima of the DoG images across scales. This is done by comparing each pixel in the DoG images to its eight neighbors at the same scale and nine corresponding neighboring pixels in each of the neighboring scales. If the pixel value is the maximum or minimum among all compared pixels, it is selected as the candidate keypoint.

4.2. Keypoint Localization

After scale space extrema are detected the SIFT algorithm discards low contrast key points and then filters out those located on edges. Scale space extrema detection produces too many keypoint candidates, some of which are unstable. The next step in the algorithm is to perform a detailed fit to the nearby data for accurate location, scale, and ratio of principal curvatures. This information allows points to be rejected that have low contrast (and are therefore sensitive to noise) or are poorly localized along an edge.

4.2.1. Interpolation of nearby data for accurate position

First, for each candidate keypoint, interpolation of nearby data is used to accurately determine its position. The approach calculates the interpolated location of the maximum, which substantially improves matching and stability. The interpolation is done using the quadratic Taylor expansion of the Difference-of-Gaussian scale-space function $D(x, y, \sigma)$ with the candidate keypoint as the origin. This Taylor expansion is given by

$$D(X) = D\frac{\partial D^T}{\partial X}X + \frac{1}{2}X^T\frac{\partial^2 D}{\partial X^2}X \quad (4)$$

Where D and its derivatives are evaluated at the candidate keypoint and $X = (x, y, \sigma)$ is the offset from this point. The location of the extremum, is determined by taking the derivative of this function with respect to x and setting it to zero. If the offset is larger than 0.5 in any dimension, then that is an indication that the extremum lies closer to another candidate keypoint. In this case, the candidate keypoint is changed and the interpolation performed instead about that point. Otherwise the offset is added to its candidate keypoint to get the interpolated estimate for the location of the extremum

4.2.2. Discarding low-contrast key points

To discard the key points with low contrast, the value of the second-order Taylor expansion $D(X)$ is computed at the offset. If this value is less than 0.03, the candidate keypoint is discarded. Otherwise it is kept, with final location and scale σ.

4.2.3. Eliminating edge responses

The DoG function will have strong responses along edges, even if the candidate keypoint is unstable to small amounts of noise. Therefore, in order to increase stability, we need to eliminate the key points that have poorly determined locations but have high edge responses.

For poorly defined peaks in the DoG function, the principal curvature across the edge would be much larger than the principal curvature along it. Finding these principal curvatures amounts to solving for the eigen values of the second-order Hessian matrix, H,

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (5)$$

The eigen values of H are proportional to the principal curvatures of D. It turns out that the ratio of the two eigen values, say α is the larger one, and β the smaller one, with ratio $r = \alpha/\beta$, is sufficient for SIFT's purposes. The trace and determinant of H are given by,

$$T_r(H) = D_{xx} + D_{yy} = \alpha + \beta \quad (6)$$

$$Det(H) = D_{xx}D_{yy} - \left(D_{xy}\right)^2 \quad (7)$$

Then

$$\frac{T_r(H)^2}{Det(H)} = \frac{(\alpha+\beta)^2}{\alpha\beta} \quad (8)$$

Since, $\alpha = r\beta$, therefore,

$$\frac{T_r(H)^2}{Det(H)} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \quad (9)$$

this depends only on the ratio of the eigen values rather than their individual values. Ratio is minimum when the eigen values are equal to each other. Therefore the higher the absolute difference between the two eigen values, which is equivalent to a higher absolute difference between the two principal curvatures of D. We use $r = 10$, which eliminates key points that have a ratio between the principal curvatures greater than 10.

4.3. Orientation Assignment

In this step, each keypoint is assigned one or more orientations based on local image gradient directions. This is the key step in achieving invariance to rotation as the keypoint descriptor can be represented relative to this orientation and therefore achieves invariance to image rotation. First, the Gaussian-smoothed image $L(x, y, \sigma)$ at the key point's scale σ is taken so that all computations are performed in a scale-invariant manner. For an image sample, $L(x, y)$ at this scale, the gradient magnitude $m(x, y)$, and orientation $\theta(x, y)$, are pre-computed using pixel differences.

$$m(x,y) = \sqrt{\left(L(x+1,y) - L(x-1,y)\right)^2 + \left(L(x,y+1) - L(x,y-1)\right)^2} \quad (10)$$

and, $\theta(x,y) = \tan^{-1}\frac{(L(x,y+1)-L(x,y-1))}{(L(x+1,y)-L(x-1,y))}$ (11)

The magnitude and direction calculations for the gradient are done for every pixel in a neighboring region around the keypoint in the Gaussian-blurred image L. An orientation histogram with 36 bins is formed, with each bin covering 10 degrees. Each sample in the neighboring window added to a histogram bin is weighted by its gradient magnitude and by a Gaussian-weighted circular window with a σ that is 1.5 times that of the scale of the keypoint. The peaks in this histogram correspond to dominant orientations. Once the histogram is filled, the orientations corresponding to the highest peak and local peaks that are within 80% of the highest peaks are assigned to the keypoint. In the case of multiple orientations being assigned, an additional keypoint is created having the same location and scale as the original keypoint for each additional orientation.

4.4. Keypoint descriptor

Previous steps found keypoint locations at particular scales and assigned orientations to them and ensured invariance to image location, scale and rotation. Now we compute descriptor vectors for these key points such that the descriptors are highly distinctive and partially invariant to the remaining variations, like illumination, 3D viewpoint, etc. The feature descriptor is computed as a set of orientation histograms on (4 x 4) pixel neighborhoods. The orientation histograms are relative to the keypoint orientation and the orientation data comes from the Gaussian image closest in scale to the key point's scale, the contribution of each pixel is weighted by the gradient magnitude, and by a Gaussian with σ equal to one half the width of the descriptor window. Histograms contain 8 bins each, and each descriptor contains a 4x4 array of 16 histograms around the keypoint. This leads to a SIFT feature vector with (4 x 4 x 8 = 128 elements). This vector is normalized to enhance invariance to changes in illumination.

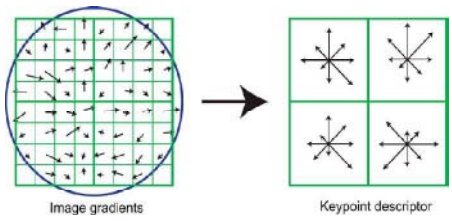

Figure 2. A 2×2 keypoint descriptor

5. Results

Results are obtained by applying the algorithm on the ear database of Indian Institute of Technology, Delhi.

5.1. Results showing SIFT key points on ear images of two different subjects

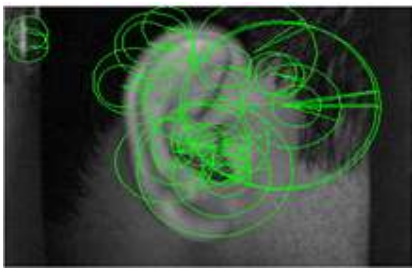

Figure 3. Key points shown on ear image of subject 1

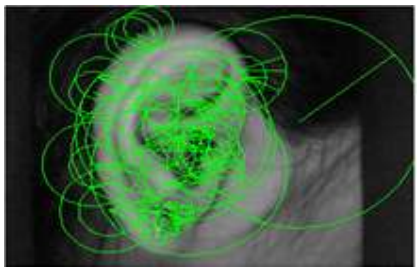

Figure 4. Key points shown on ear image of subject 2

5.2. Results of using SIFT on ear for matching under controlled environment

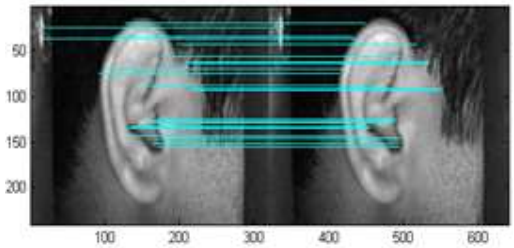

Figure 5. Keypoint matching of same ear of subject 1 (41 matches found out of 43)

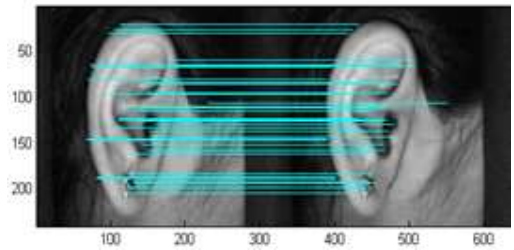

Figure 6. Keypoint matching of same ear of subject 2 (77 matches found out of 81)

5.3. Results of using SIFT on ear for matching after rotation

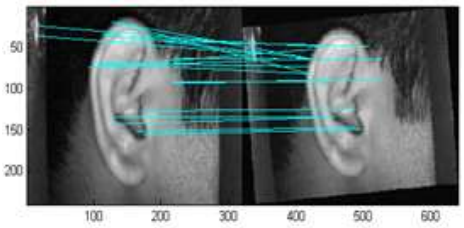

Figure 7. Keypoint matching after rotation by 5 degrees in counterclockwise direction (27 matches found out of 43)

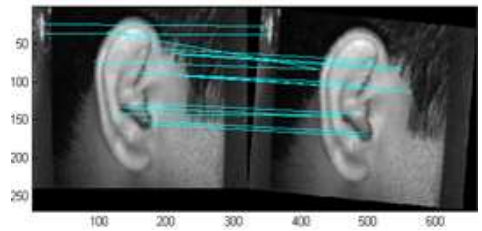

Figure 8. Keypoint matching after rotation by 5 degrees in clockwise direction (25 matches found out of 43)

5.4. Results of using SIFT on ear for matching after changing the scale

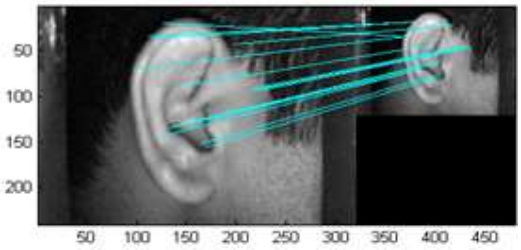

Figure 9. Keypoint matching after reducing size of ear image of subject 1 (23 matches found out of 43)

5.5. Results of using SIFT for matching ears of different subjects

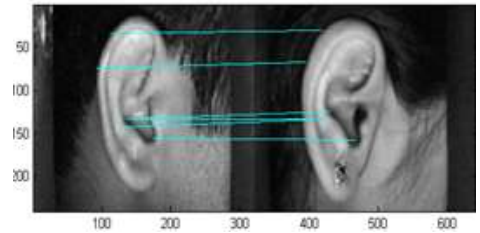

Figure 10. Keypoint matching of different subjects (only 7 matches found)

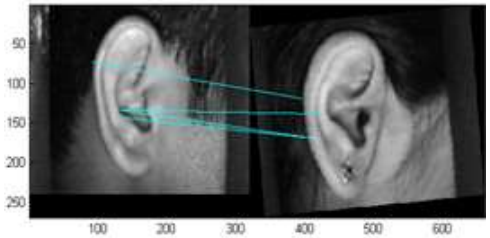

Figure 11. Keypoint matching of different subjects after rotation (only 4 matches found)

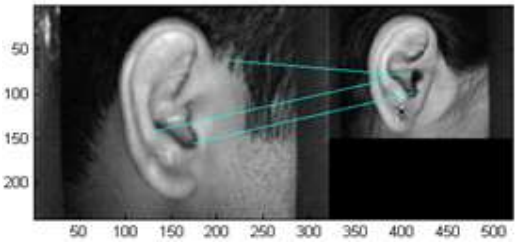

Figure 12. Keypoint matching of different subjects after changing the scale (only 4 matches found)

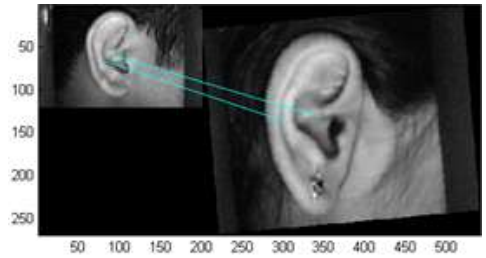

Figure 13. Keypoint matching of different subjects after changing scale and rotation (only 2 matches found)

6. Conclusion

From several experiments conducted we conclude that the SIFT algorithm gives good results for matching under both controlled and strained environment. It is also robust to changes in scale and rotation for the ear biometric, but the results after changing illumination is not so good. It gives good results for small occlusion but not for large occlusion. It is an efficient algorithm for recognition and detection purpose for small data base but not so good (with respect to computational speed) for large database.

7. Future Scope

In SIFT edges are poorly defined and usually hard to detect, but there are still large numbers of key points can be extracted from typical images. Also sometimes the images are too smooth to find that many SIFT features for a matching, and in that case a small image could be unrecognized from the training images. The recognition performance could be improved by adding new SIFT features or varying feature sizes and offsets.

References

[1] A. Iannarelli, Ear Identification, Forensic Identification Series, Paramont Publishing Company, Fremont, California, 1989.

[2] Mark Burge, Wilhelm Burger, "Ear Biometrics in Computer Vision", 15th International conference On Pattern Recognition (ICPR), Volume 2, pp.2822, 2000.

[3] David G. Lowe, "Distinctive Image Features from Scale-Invariant Key points", International Journal of Computer Vision, vol.60 (2), pp.91-110, 2004.

[4] David G. Lowe, " Local feature view clustering for 3D object recognition", IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, pp. 682-688, 2001.

[5] Cong Geng, Xudong Jiang, "Face Recognition Using SIFT Features", ICIP, IEEE, 2009.

[6] David G. Lowe, "Object Recognition from Local Scale-Invariant Features", The Proceedings of the Seventh IEEE International Conference on Computer Vision, vol.2, pp.1150-1157, 1999.

[7] Chunmei Zhang, Zhihui Gong, Lei Sun, "Improved SIFT Feature applied in image Matching", Computer Engineering and Application, Vol.2, pp.95-97, 2008.

[8] Xiaomin Liu, Peihua Li, "An Iris Recognition Approach with SIFT Descriptors", ICIC, LNAI 6839, pp.427-434, Springer-Verlag Berlin Heidelberg, 2012.

[9] Andrea Lingua, Davide Marenchino, Francesco Nex, "Performance Analysis of the SIFT Operator for Automatic Feature Extraction and Matching in Photogrammetric Applications", pp.3745-3766, Sensors, 2009.

Author Profile

Himanshu Maurya received the B.E. degree in Electronics and Communication Engineering from Thakral college of Technology in 2011, and presently pursuing M.Tech in Communication System from National Institute of Technology, Patna, India.

Shikha Maurya received the B.Tech degree in Electronics and Communication Engineering from B. N. College of Engineering of Technology in 2012, and presently pursuing M.Tech in digital Communication from ABVIIITM, Gwalior, India.