ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

An approach for identification using ear and face biometrics employing score based image fusion and SIFT feature detection in surveillance system

Akash Pal [1], Avinandan Sau[2], Debolina Das[3], Sukanta Shrama[4], Madhuchhanda Dasgupta [5]

Student, Computer Science & Engineering, JIS College of Engineering, Kalynai, India[1,2,3,4]

Assistant Professor, Computer Science & Engineering, JIS College of Engineering, Kalynai, India[5]

Abstract: Biometric surveillance is any technology which measures and analyzes human physical and/or behavioural characteristics for authentication, identification, or screening purposes. Identification is an essential part of our lives. Identification of authentic candidate is essential in E-commerce, in keeping track of criminals, in airport and railway surveillance and many more aspects of the modern world. However identification of a person can be challenging especially when the person is not cooperating. This leads to classify the identification techniques broadly in to two categories: Passive and Active. In this current study an approach has been proposed combining the ear and face biometrics for the purpose of identification of a person using SIFT (Scale Invariant Feature Transform). The proposed multi biometric system achieves a recognition accuracy of 99.9958%.

Keywords: SIFT, biometric fusion, ear-face biometric.

I. INTRODUCTION

Biometric: Biometrics are our most unique physical (and behavioural) features that can be practically sensed by devices and interpreted by computers so that they may be used as proxies of our physical selves in the digital realm. In this way we can bond digital data to our identity with permanency, consistency, and un-ambiguity, and retrieve that data using computers in a rapid and automated fashion.

There are two types of biometric system

1. Unimodalbiometric: In unimodal biometric system only one biometric trait is used for the purpose of identification of a person.
2. Multimodal biometric: In multi-modal biometric system more than one biometric trait is used for the purpose of identification.

However unimodal biometric system has certain disadvantages due to which the multimodal biometric systems are preferred. There are two stages of a biometric system: Enrolment and Authentication. Enrolment is the process of the creation of users' sample based on certain biometric traits and storing it in the user database. During the Authentication phase the identity of a user is verified by obtaining the biometric traits and comparing with the stored one. The user is accepted if it is a appropriate match.

II. PURPOSE OF THIS WORK

This research work has been carried out to accomplish the following objectives:

1. SIFT as the feature extraction algorithm.
2. Propose a new multi-biometric trait combining side-face and ear for a surveillance system.
3. Obtaining the accuracy of this proposed system.

III. PROPOSED SYSTEM

In surveillance system, the objective is to identity an individual(s) successfully whose records are present in the database. The biometric traits have to be acquired passively. If only the side face is captured then from this side face two biometric traits can be obtained the ear and a portion of the face. Since ear as a biometric system is not effective combining with the face-portion could give better result. In this paper, a database of side-faces is used to verify the effectiveness and the accuracy of the proposed system.

IV. STEPS

1. Image acquisition
2. Region of Interest extraction
2. Feature Extraction
3. Matching
4. Score level Fusion
5. Performance analysis

V. IMAGE ACQUISITION

For this purpose the CVL (Computer Vision Laboratory) Face database has been used. It contains facial images of 114 subjects, with 7 images for each person. Each of the images was taken with a Sony Digital Mavica under uniform illumination, no flash and with projection screen in the background. All the images are of resolution 640*480 pixels in jpeg format.

All the subjects were mostly male (around 90%) around 18 (pupils and some professors). For this work, only one image (i.e. the side face) per subject is used. Out of 114 subjects only 109 subjects were considered as 5 of had occluded ear.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

Table 1: Side faces from CVL database

VI. REGION OF INTEREST EXTRACTION

The face and the ear of each person can be extracted automatically using techniques such as template matching. However, here the face and the ear of each person are extracted manually by cropping.

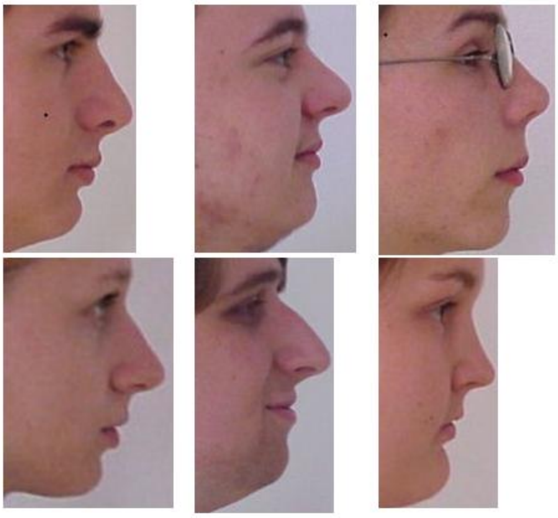

Table 2: Face profile images

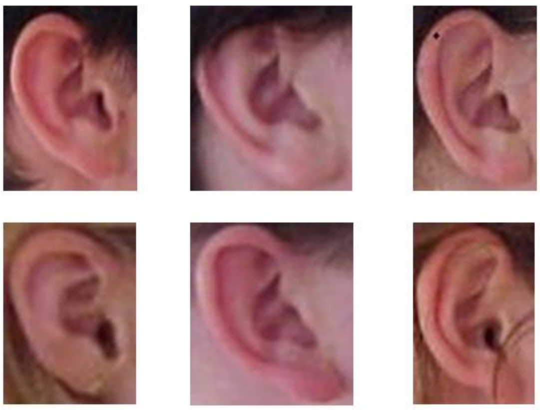

Table 3: Ear images

VII. FEATURE EXTRACTION

For the purpose of feature extraction the SIFT (Scale Invariant Feature Transform) algorithm proposed by D.Lowe [3] has been used.

A. SIFT Descriptor

The scale invariant feature transform, called SIFT [6] descriptor, has been proposed by and proved to be invariant to image rotation, scaling, translation, partly illumination changes. Following are the major stages of computation used to generate the set of image features

B. Scale Space extrema detection

The first stage of computation is to create a scale space of images. This is done by constructing a set of progressively Gaussian blurred images with increasing values of sigma. Then the difference between pairs of Gaussian is taken to obtain a Difference of Gaussian (DOG) which is similar to the function Laplacian of Gaussian (LOG) to obtain potential locations for finding features. The image is then sub-sampled (i.e. $1/4^{th}$ resolution of lower octave) to obtain the next octave and the same process is repeated to obtain DOG pyramid.

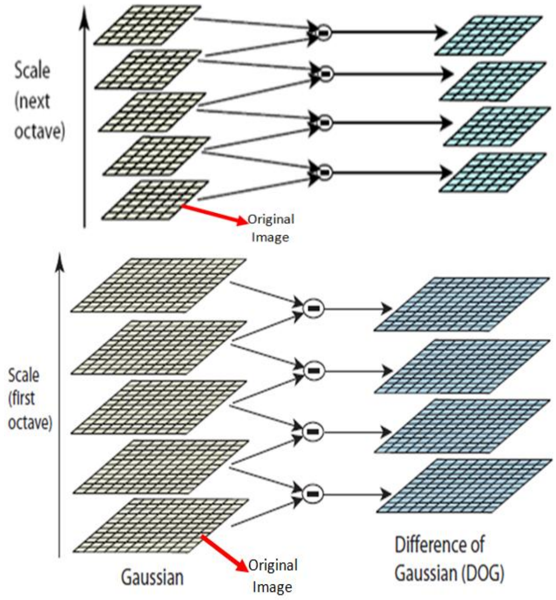

Figure 1: Operations within same octaves (same scale)

C. Keypoint localization

Accurately locates the feature key points by comparing a pixel (X) with 26 pixels in current scale and adjacent scales (Green Circles). The pixel (X) is selected if it is larger/smaller than all 26 pixels. There are still a lot of points; some of them are not good enough. The locations of keypoints may be not accurate. Eliminating edge points, keypoints are selected from the extrema based on measures of their stability.

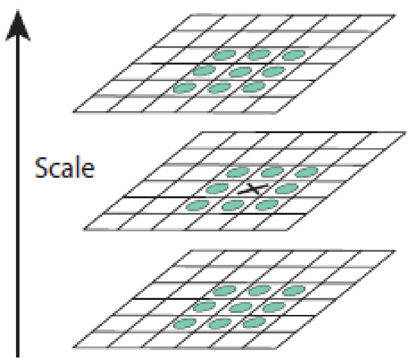

Figure 2: Operations between different octaves (different scale)

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

D. Orientation assignment

This step assigns orientation to each keypoint, the keypoint descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation. It computes magnitude and orientation on the Gaussian smoothed images. An orientation histogram is formed from the gradient orientations of sample points within a region around the keypoint. Peaks in the orientation histogram correspond to dominant directions of local gradients. The highest peak in the histogram is detected, and then any other local peak that is within 80% of the highest peak is used to also create a keypoint with that orientation. One or more orientations are assigned to each keypoint location based on local image gradient directions.

36 bins (10 degree each)

if 1<=orientation<=10 ;	**bin 1**
else if 11<= orientation <=20 ;	**bin 2**
else if 21<= orientation <=30 ;	**bin 3**
.	
.	
.	
else if 351<= orientation <=360 ;	**bin 36**

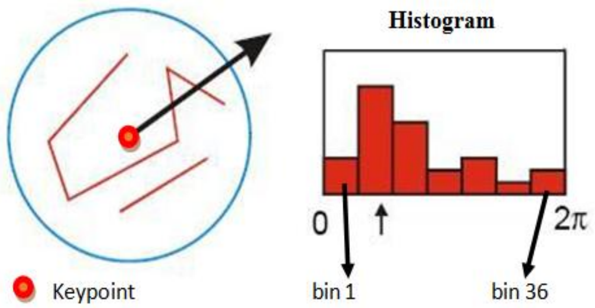

Figure 3: Orientation assignment

E. Keypoint descriptor

This step describes the key point as a high dimensional vector. The local image gradients are measured at the selected scale in the region around each keypoint. It computes relative orientation and magnitude in a 16x16 neighborhood around each key point.

It forms weighted histogram (8 bin) for 4x4 regions. Finally it concatenates 16 histograms in one long vector of 128 dimensions.

These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination. This approach has been named the Scale Invariant Feature Transform (SIFT), as it transforms image data into scale invariant coordinates relative to local features.

An important aspect of this approach is that it generates large number of features that densely cover the image over the full range of scales and locations.

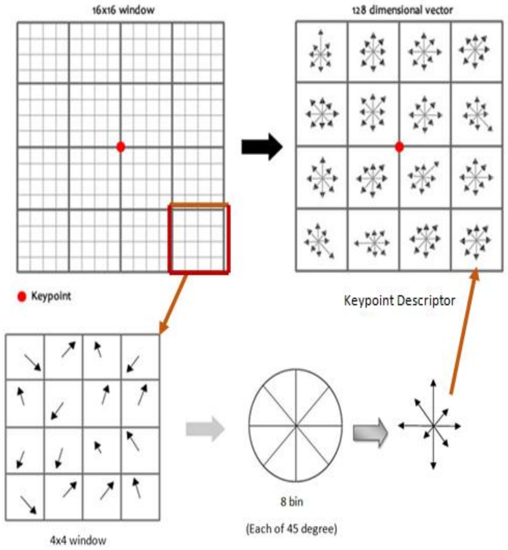

Figure 4:The Key point Descriptor

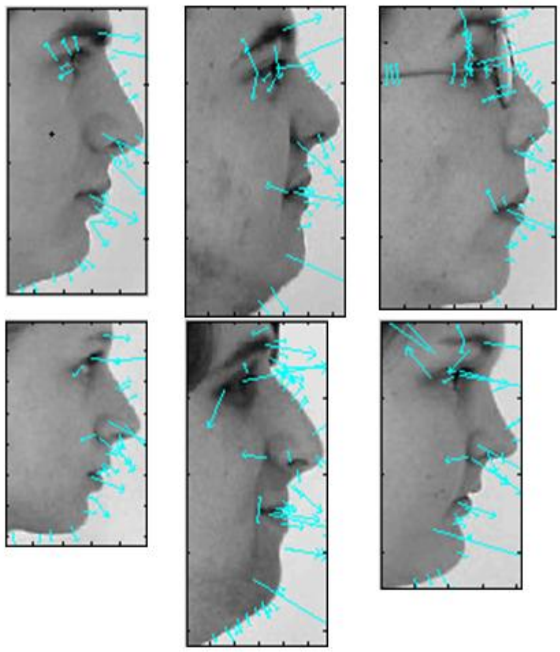

Table 4: Face profile images with the SIFT features

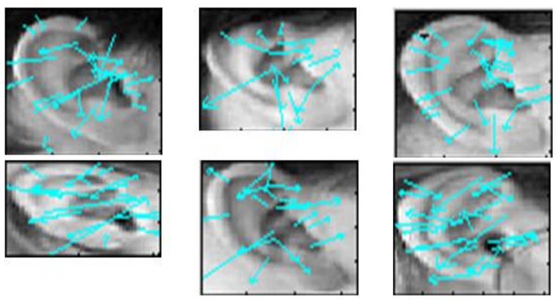

Table 3: Ear images with the SIFT features

VIII. STEPS OF MATCHING

1. Before matching two images, the feature descriptors of the two images are obtained. Now for each descriptor in the first image, a match is selected in the second image.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

IMAGE 1

Keypoints	Keypoint descriptor
1	f1
2	f2
3	f3
. . .	. . .
m	fm

IMAGE 2

Keypoints	Keypoint descriptor
1	f1
2	f2
3	f3
. . .	. . .
n	fn

Figure 5: Step 1 of matching

2.For obtaining the match of the first keypoint in the first image. The dot product is calculated between the first keypoint descriptor in the first imagewith the other keypoint descriptors in the second image and is denoted by v1,...,vn. After calculating the dot product, the keypoints are sorted in ascending order based on the dot product. This is done for every keypoint in the first image.

For every Keypoint in IMAGE 1

Keypoints (IMAGE 1)	distance/ dot product (sorted)
2	v2
3	v3
1	v1
. . .	. . .
n	vn

Figure 6: Step 2 of matching

3. In this step, the ratio of the two smallest values is taken. If this value is less than the distance ratio then the first keypoint in the sorted list is taken to be the matching keypoint for the first keypoint in first image.

$$v2/v3 < distance\_ratio$$

For every Keypoint in IMAGE 1

Keypoints (IMAGE 1)	distance/ dot product (sorted)
2	v2
3	v3
1	v1
. . .	. . .
n	vn

Keypoints (IMAGE 1)	Keypoint (IMAGE 2)
1	2
2	
3	
. . .	. . .
m	

Figure 7: Step 3 of matching

4. This entire process is repeated for obtaining the matching keypoints between the two images.
5. Finally the two images are appended side-by-side and then a straight line is drawn between the matched keypoints. Also the matching score between the images is given.

IX. MATCHING

A. True match

When an image is matched against itself, it is called a true match. In this case the number of keypoints generated is large.

If the first face-portion image is compared with itself then 36 match points were found. Figure below shows the result of matching.

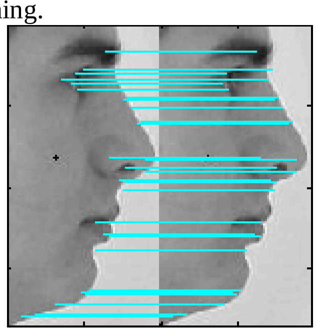

Figure 8: True match of first face-profile image

If the first ear image is compared with itself then 27 match points were found. Figure below shows the result of matching.

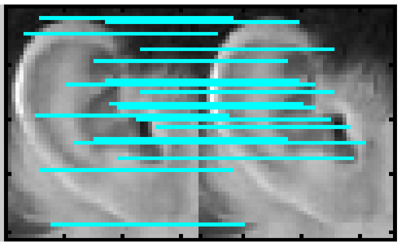

Figure 9: True match of first ear image

B. False Match

When an image is matched against an imposter image, it is called a false match. In this case the number of keypoints generated is very less. If the first face-portion image is compared with the second face-portion image then 3 match points were found. Figure below shows the result of matching.

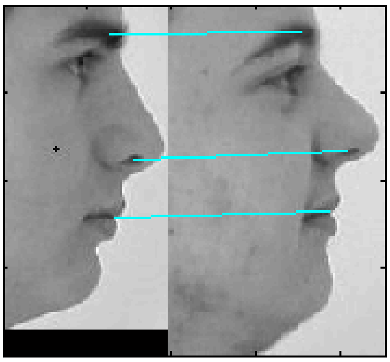

Figure 10: False match of first face-profile image

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

If the first ear image is compared with the second ear image then 0 match points were found. Figure below shows the result of matching.

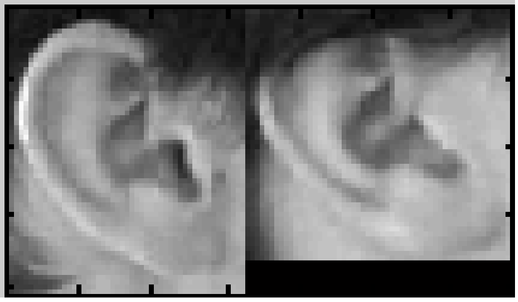

Figure 11: False match of first ear image

X. PERFORMANCE ANALYSIS

FRR (False Rejection Rate) is the process of falsely rejecting a genuine user.

$$\text{FRR (False Rejection Rate)} = \frac{\text{Number of false Rejections}}{\text{Total number of authentic attempts}}$$

FAR (False Acceptance Rate) is the process of falsely accepting an imposter.

$$\text{FAR (False Acceptance Rate)} = \frac{\text{Number of false Acceptance}}{\text{Total number of imposter attempts}}$$

EER (Equal Error Rate) is defined as the point where FAR is equal to FRR.

$$\text{EER (Equal Error Rate)} = \frac{\text{FAR} + \text{FRR}}{2}$$

$$\text{Accuracy} = 100 - \text{EER}$$

Receiver Operating Characteristic (ROC) curve:

This curve is used tosummarize the performance of a biometric verification system. An ROC curveplots, parametrically as a function of the decision threshold, the percentage of impostorattempts accepted (i.e. false acceptance rate (FAR)) on the x-axis, againstthe percentage of genuine attempts accepted (i.e. 1 - false rejection rate (FRR)) on the y-axis. The ROC curve is threshold independent, allowing performancecomparison of different systems under similar conditions.

Detection Error Trade-off (DET) curve

In the case of biometric systems,the DET curve is often preferred to the ROC curve. Indeed, the DET curve plotserror rates on both axes (FAR on the x-axis against FRR on the y-axis) usingnormal deviate scale what spreads out the plot and distinguishes differentwell-performing systems more clearly.

109 numbers of images are used from the CVL database. For finding **FRR** each image is compared against itself, which results in 109 authentic attempts into the proposed system. For finding **FAR** each image is compared with the other 108 images, which results in 11772 (109x109-109) imposter attempts into the proposed system. The FAR and FRR can be combined together into matching score matrix.

1	22	0	0	1	1	0	0	0	0	0
2	0	21	0	0	0	1	0	0	0	1
3	0	0	27	1	1	2	0	1	0	3
4	0	0	1	12	1	4	1	1	0	0
5	0	0	1	5	19	0	0	1	0	0
6	1	0	1	4	3	22	5	3	0	2
7	0	0	1	1	0	5	22	5	0	0
8	0	0	0	0	1	0	2	16	1	0
9	0	0	0	0	0	0	0	0	30	0
10	0	0	1	0	0	1	0	2	0	32

Figure 12: Snapshot of the matching scores of 10 users.

In this matrix the diagonal scores represents the FRR and the rest of the scores except the diagonal elements represents the FAR.

In case of similarity match, if the matching score is more than the threshold then it is a match or else not.

Similarity Match: If
Match Score > Threshold
then
Match
else
Not Match

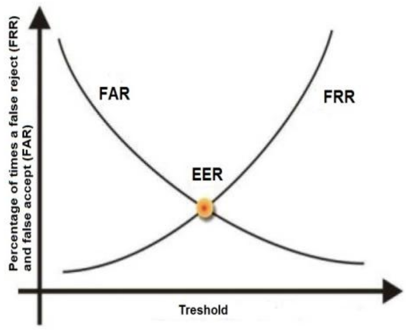

Figure 13: Ideal FAR/FRR vs Threshiold curve

XI. FUSION TECHNIQUES

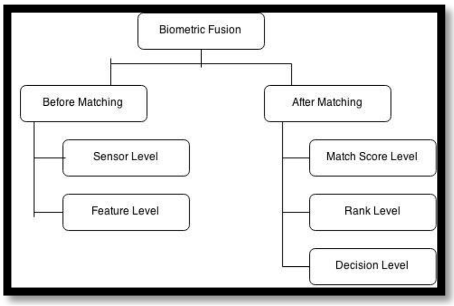

Figure 14: Levels of fusion

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 4, April 2015

For the proposed system, the score level fusion technique has been adopted. The matching score matrix of ear and face-portion are combined together into a single matching score matrix using sum rule. This matching score matrix is used for performance analysis.[5]

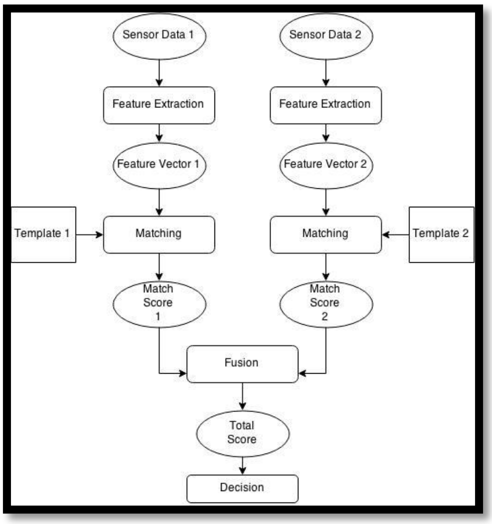

Figure 15: Match score level fusion

XII. CURVES

After score level fusion

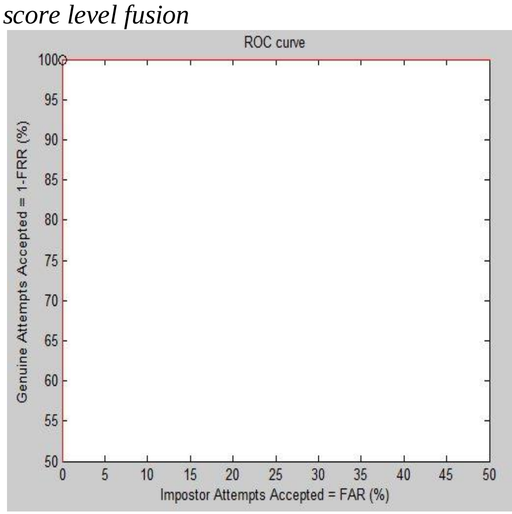

Figure 16: ROC curve after score level fusion

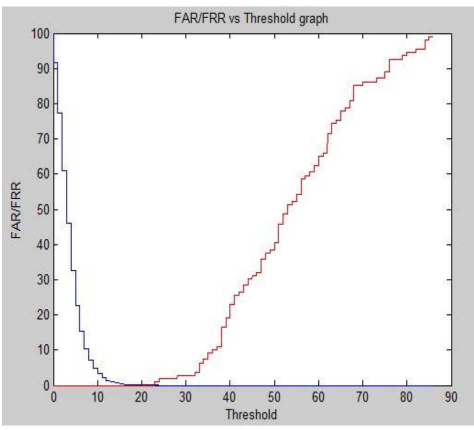

Figure 17: FAR/FRR vs Threshold curve after score level fusion

XIII. RESULT

A. Ear images
FAR: 0.7985
FRR: 0
EER: 0.3993
Accuracy = 99.6007 %

B. Face portion
FAR: 0.5607
FRR: 0
EER: 0.6626
Accuracy = 99.3374 %

C. After fusion
FAR: 0.0085
FRR: 0
EER: 0.0042
Accuracy = 99.9958 %

	Rahmanet al[8]	Yuan el al.[9]	Panet al.[10]	Xu et al.[7]	Youssefet al.[11]	Proposed Technique
Technique Applied	PCA	FSLDA	KFDA	FSLDA	LBP	SIFT
Database Used	UND	USTB	USTB	USTB	UND	CVL
Subject Number	18	79	79	42	102	109
Face Profile Results	88.88%	-	93.46%	88.10%	96.76%	99.3374%
Ear Results	77.77%	-	91.77%	94.05%	96.95%	99.6007%
Fusion Results	94.44%	96.20%	96.84%	97.62%	97.98%	99.9958%

Table 5: Comparison between various techniques

XIV. CONCLUSION

In this paper a new multi-biometric system has been proposed mainly for the surveillance system whereby the side face is used to extract the face-profile and ear. In this work, the process of extraction has been done manually, however this could be automated in future using techniques such as template matching. The SIFT algorithm is the best feature extraction algorithm which could be used in biometric security as it gives a recognition accuracy of around 99%. The SIFT algorithm would provide similar results independent of the database being used. The proposed multi-biometric system gives higher recognition accuracyon combining the unimodal biometric traits of face-profile and ear using simple summation technique of the matching scores.

Another important thing to note is the value of FRR which is zero for both the unimodal and multimodal trait. This implies that using the SIFT algorithm completely removes the possibility of accepting an imposter. However the value of FAR is not exactly zero, so there is even a small possibility that a genuine user may be rejected. The value of FRR makes the proposed system perfect for use in surveillance system. As the objective in surveillance system is to correctly identify an individual whose records are stored.

REFERENCES

[1]. Wang, Yunhong, Tieniu Tan, and Anil K. Jain. "Combining face and iris biometrics for identity verification." Audio-and Video-Based Biometric Person Authentication. Springer Berlin Heidelberg, 2003.

[2]. Karanwal, Shekhar, Davendra Kumar, and RohitMaurya. "Fusion of fingerprint and face by using DWT and SIFT." Int J ComputAppl 2.5 (2010): 0975-8887.

[3]. Lowe, David G. "Distinctive image features from scale-invariant keypoints."International journal of computer vision 60.2 (2004): 91-110.
[4]. Bagal, Ms VL, N. B. Sambre, and P. Malathi. "Identification of Person by fusion and using Scale Invariant Feature Transform."
[5]. Verma, Shalini, and R. K. Singh. "Multimodal Biometrics Information Fusion for Efficient Recognition using Weighted Method."
[6]. Cummings, Alastair H., Mark S. Nixon, and John N. Carter. "A novel ray analogy for enrolment of ear biometrics." Biometrics:
[7]. Xu, X. and Mu, Z.-C., "Multimodal recognition based on fusion of ear and profile face," in [Proc. of the 4th InternationalConference on Image and Graphics ICIG], 598–603 (2007).
[8]. Rahman,M.M.Ishikawa ,Proposing a passive biometric system for robotic vision," in [Proc. of the 10th International Symposium on Artificial Life and Robotics (AROB)], (Feb 2005)
[9]. Yuan, L., Mu, Z.-C., and Liu, Y., "Multimodal recognition using face profile and ear," in [Proc. of the IEEE International Symposium on Systems and Control in Aerospace and Astronautics ISSCAA], 887–891 (2006).
[10]. Pan, X., Cao, Y., Xu, X., Lu, Y., and Zhao, Y., "Ear and face based multimodal recognition based on KFDA," in[Proc. of the International Conference on Audio, Language and Image Processing ICALIP], 965–969 (2008).
[11]. Youssef, Iman S., et al. "Multimodal biometrics system based on face profile and ear." SPIE Defense+ Security. International Society for Optics and Photonics, 2014.

BIOGRAPHIES

Mr. Akash Pal is a student of final year B.Tech in Computer Science & Engineering Department in JIS College of Engineering, West Bengal, India. He has attended various national and international conferences. He has been awarded as the beast researcher by the Computer Science & Engineering Department in JIS College of Engineering. His research interest includes image processing, biometric security, ecommerce security, and robotics.

Mr. Avinandan Sau is a student of final year B.Tech in Computer Science & Engineering Department in JIS College of Engineering, West Bengal, India. His research interest includes image processing, biometric security, cyber security, and robotics.

Mr. Sukanta Sharma is a student of final year B.Tech in Computer Science & Engineering Department in JIS College of Engineering, West Bengal, India. His research interest includes image processing, biometric security.

Miss. Debolina Das is a student of final year B.Tech in Computer Science & Engineering Department in JIS College of Engineering, West Bengal, India. Her research interest includes Image Processing, Cyber Security.

Madhuchhanda Dasgupta has received her M.Tech (Computer Science & Engg.) from Kalyani Govt. Engineering College in 2008. She is currently working as an Assistant Professor in Computer Science & Engineering Department in JIS College of Engineering, West Bengal, India. Her area of interest includes image processing, Pattern Recognition and Automata.