RESEARCH ARTICLE OPEN ACCESS

Hybrid Recognition System under Feature Selection and Fusion

Eng. Ali M Mayya [1], Dr. Mariam M Saii [2]
PhD Student [1], Professor Assistance [2]
Department of Computer Engineering
University of Tishreen, Lattakia
Syria

ABSTRACT

The study suggests a hybrid human recognition system based on face ear and palm print images. The aim of our study is to show the importance of biometric fusion for enhancement of the recognition rate. The system takes face images and segment them into face and ear. Then face, ear and palm images are extracted and fused. The FFBPNN followed by mahalanobis classifier is used in classification stage. The Experiments is applied on 5 databases with different illumination and pose variations and the best result obtained from the face-ear-palm fusion features with 97.5% recognition rate compared with 94.33% for face-ear fusion.

Keywords:- Fusion, Biometrics, Human Recognition, Palmprint, Face, Ear.

I. INTRODUCTION

Many of recent studies focus on biometric fusion like face-ear, face-iris, iris-palm etc., but all of these researches didn't care about the big vector size which may increase the processing time without significant enhancement of performance, and didn't discuss the effect of different biometric fusion with different datasets.

Islam [1] depended on fusion at feature and score-level for face and ear images. 3D local features had been extracted and using the iterated closet point (ICP) and "sum rule", the fusion step was done. The Islam's system depended on a database consisting of 326 individual. The system got 98.4% recognition rate for feature level fusion and 99% for score level fusion. The system took in account illumination and pose variation and occlusion by hair or earrings but the ICP algorithm failed in case of large pose variations.

GARCIA [2], at 2011, used right and left face images for extraction the features by different algorithms (PCA for feature extraction, FFBPNN for classification). The fusion part was done by concatenation at feature level and by majority voting at score level. The system achieved 91.8% recognition rate at data level fusion and 98.97% at feature level fusion

. In 2015, Pal [3] introduced a face-ear system to recognize people. He extracted ear from side face image manually, and extracted features based on SIFT algorithm which extract variation-resistance features. By using minimum distance between training and test samples, the classification decision was made. Pal used 228 images taken from 114 individuals (1 image for training and 1 for test). The system didn't take in account any illumination or pose variations or partial occlusion by hair or earrings. As a result, the system got 99.9% recognition rate. In 2016, Lei [4] used Log Gabor filter to extract face and ear features. By using Kernel Fisher discriminant analysis KFDA, the best features were selected from the fused face and ear features. 339 images form 113 individuals (3 images per one). System didn't define test images' nature or count while the recognition rate was 91.15% for face, 93.81% for ear and 99.12% in case of fusion. The system took in account illumination and pose variations but didn't take care about occlusion. The research illustrated that recognition rate could decreased by 0.88% in case of face variations. Shams [7] introduced a three biometric system consists of face, iris and fingerprint. They used hough transform for iris segmentation. At feature extraction stage, they used LBPV method, and at the classification stage they used LVQ classifier with 99.09% recognition rate under a 50 person dataset. Recently, VALARMATHY [8] used three biometrics which are face, palm and iris, they used the LBP technique for feature extraction and K-nn for classification. The fusion is done at feature level using concatenation and the system achieved 99% recognition rate but the FAR was high.

Most of studies in fusion field did not take in account time required for fusion phase due to large feature vectors or redundant information. Our suggested system focuses on selecting the promised features of the feature vector.

II. MATERIAL AND METHODS

The suggested system consists of multiple stages from preprocessing until classification. Figure 1 illustrates these stages.

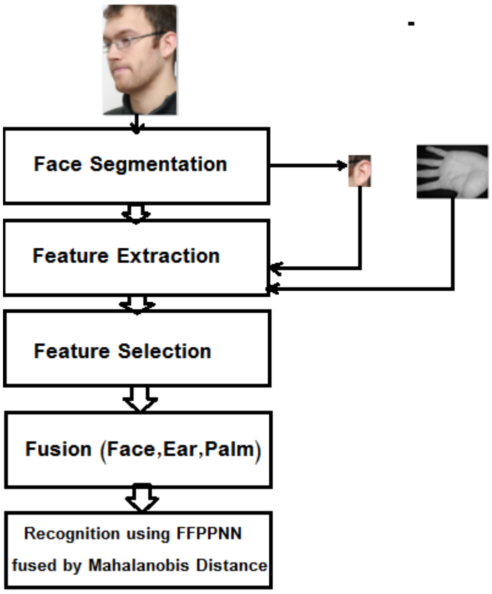

Fig1. Block Diagram of the Proposed System

2.1 Preprocessing (Illumination Correction):

The suggested approach for correction the illumination starts with resizing the image into specific size to unite the next steps. The histogram of the resized image is calculated and the very high gray levels is detected and encountered in variable K. After that, the illumination is corrected via modified Gamma correction method given as illustrated in equation1.

$$s(x,y) = \begin{cases} f(x,y)^{0.7} & if\ numpixl < \mu(x,y) * (m * n/10000) \\ f(x,y) & otherwise \end{cases}$$

Where s(x,y) is the corrected gray level, f(x,y) is the original gray level, $\mu(x,y)$ is the mean value of the original image, m and n are the image dimensions, 10000 is the scope of image dimensions. Gamma value is optional and for our algorithm we chose 0.7. So, equation 1 supposes that if K is less than $\mu(x,y) * m * n/1000$ then the illumination is low and must be corrected by gamma=0.7 (logarithm effect), on the other hand, if k is more than D the illumination is high and must be modified via 1/gamma (exponential effect). In other cases where the k variable has other values, the image is considered to be normal and no correction is made.

Figure 2 introduces the illumination correction functions and two examples of corrections.

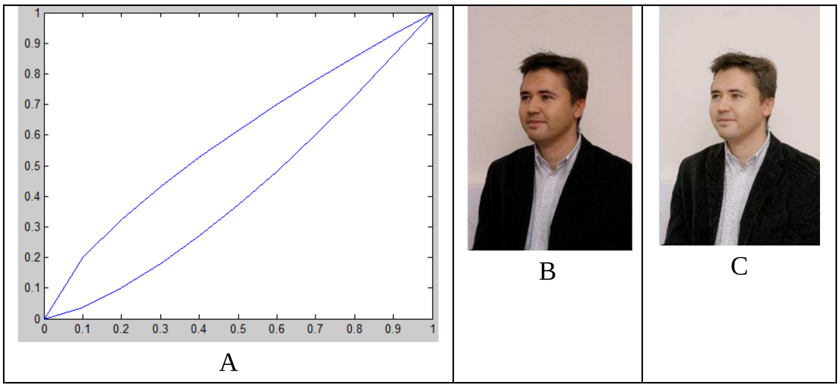

Fig 2. Illumination Correction: A: Illumination correction function, (B): original images, (C): corrected images.

2.2 Face Region Extraction:

This step aims to extract the basic face region. The approach of four simple steps which are the color segmentation of face-like regions, morphological closing operation with disk structural element, filling holes operation and elimination of small regions which are not faces.

The color segmentation is depends on the skin color range which is (above 170 for Red channel, 90-150 for Green, 90-110 for blue). The image is then transformed into binary by low threshold 0.2 and the closing morphological operation is done by a 5 radius disk structural element. The closed image is filled using the filling holes morphological operation. The last step eliminates regions with areas less than specified value (Non faces regions). Figure 3 illustrates a face region extraction example.

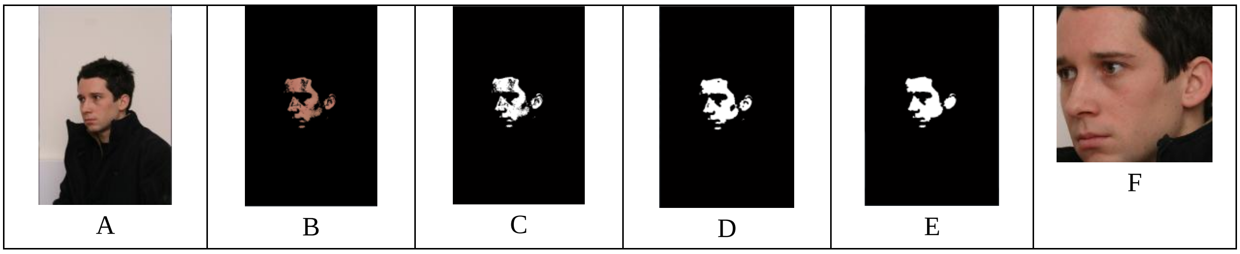

Fig3. Face Extraction Steps: A. original image, B. Color Segmentation, C. Thresholding, D. Closing, E. Filling holes and eliminate small regions, F. Face Region.

2.3 Ear Region Extraction:

The input for this stage is the cropped face region from previous stage and the output is the ear region using the following steps:

- The mean value of face image is calculated and the image is transformed into binary formula using this value. The binary image is processed by eliminating the small pixels using the "open" morphological operation to obtain the regions with area less than quarter of the maximum area. In the result image, the regions which match the following two conditions are detected while the others are eliminated. First condition is that region must have orientation between -20 and 20, and the second is that region must have extent between 0.4 and 6 and regulation between 0.5 and 4 (eye shape has rectangular extension and almost 0 orientation). The result is shown in figure (4.C).

- pose normalization: this step aims to make the face pose always to left. The approach for face normalization calculates the area in each half of the filled image. Then if the areas of components on right half are bigger than those on the left, the image must be flipped horizontally; otherwise, the pose is maintained figure (4.D).

- Left_face Elimination or Ear localization: This aims to delete the left half of face in which the ear region doesn't exist. To eliminate the left half, we detect the eye regions and eliminate the face region with x coordinates less than eye's coordinates as shown in figure 4.D.

- Ear Detection: this step depends on color segmentation to detect which uses the equation 2.

$$ER(x,y) = \begin{cases} 1 & \text{if } R(x,y)/G(x,y) \geq Th1 \ \& \ R(x,y)/G(x,y) < Th2 \ \& \ orient > 10 \ \& \ orient < 90 \\ 0 & otherwise \end{cases} \quad (2)$$

Where ER(x,y) is the extracted ear, R(x,y) and G(x,y) are the red and green channels. Th1 and Th2 are color thresholds (in our system we chose th1=1.8, th2=2.8). Figure 4.E illustrates ear region.

- The final step is to extract the ear region from face image using coordinates of the ear region obtained from previous step.

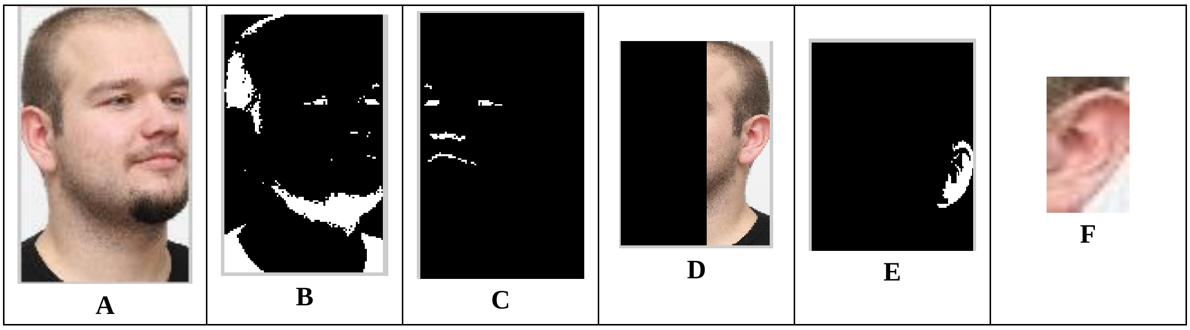

Fig 4. Ear Extraction Steps: A. Extracted Face, B. Thresholding, C:Opened ,Filled and Flipped image, D. Ear localization, E. Ear detection, F. Ear extraction.

2.4 Feature Extraction:

The features extraction phase aims to obtain the feature vector which represents the important information of image. In our system, we suppose using Wavelet 2D approximation with DB1 function which is better than Haar for classification and recognition.

. To obtain the approximation features, the LPF is applied twice and the result coefficients have quarter size of the original image (this minimize the redundant information). For more minimization, wavelet transform is applied on the approximation coefficients again to obtain approximation of level 2. The last step in feature extraction is the reshaping of approximation matrixes of face and ear into vectors of one row and M/4*N/4 columns.

The second feature of face are geometrics features, which are illustrated in figure 5.

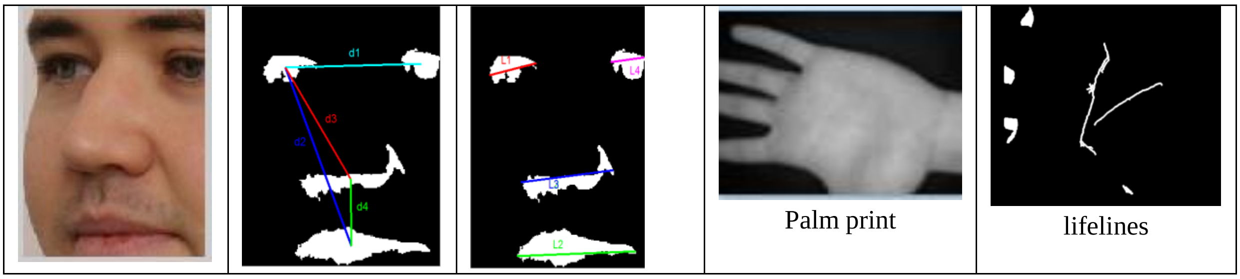

Fig 5. Face Feature Extraction Method

The features of wavelet approximation and the geometric features are fused to consist the face features. While the palm print wavelet features are fused with lifelines feature to consist the palmprint features.

2.5 Feature Selection:

Until now, the feature vector is still big (for our system 83000 samples) and will require more time for next processing steps and more size for storage. So, we introduce an improved approach from [7] for feature selection to minimize features by choose the most promised features. Figure 6 describes our modified suggested system. Our modification in the output of fuzzy system and inner coefficients.

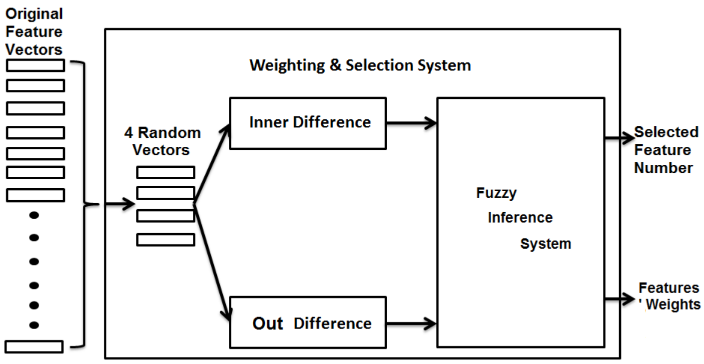

Fig 6. Suggested Weighting and Selection Model

The Weighting and selection algorithm is described below:

1- Choose N random vectors of 4 different persons (Class1, Class2, Class3, Class4) each of them contains 4 different samples F1, F2, F3,F4 for the same vector. So, we have 4*N samples of N classes. (i.e $F1_{3,3}$ is the 3rd sample of class1 where i is the class id, and j is the sample id into class).

2- Calculate the inner difference at the same class to obtain N difference value described in equation 5.

$$InDiff f_{i,j}^{k} = \sum_{i=1}^{N} \sum_{j=1}^{4} \left|F_{i,j}^{k} - F_{i,j+m}^{k}\right| \qquad (5)$$

$where\ k = 1,2,,3,4, \dots .N\ , (j+m) \leq N$

3- Calculate the difference between each two classes to obtain out difference values (Divergence) described in equation 6.

$$OutDiff f_{i,j}^{k} = \sum_{i=1}^{N} \sum_{j=1}^{K} \left|F_{i,j}^{k} - F_{i,j}^{k+1}\right| \qquad (6)$$

$where\ k = 1,2,,3, \dots, N\ , (j+m) \leq N$

4- Calculate the minimum inner difference average defined in equation 7.

$$InDiff f_{i,1}^{k} = \min\left(\frac{1}{N} \sum_{j=1}^{N} InDiff f_{i,j}^{k}\right) \qquad (7)$$

5- Calculate the maximum out difference average defined in equation 8.

$$OutDiff f_{i,1}^{k} = \max\left(\frac{1}{N} \sum_{j=1}^{N} OutDiff f_{i,j}^{k}\right) \qquad (8)$$

6- Build a fuzzy inference system with two inputs which are inner and out difference ($diff_{in}{}_{i,j}^{k}$ and $diff_{out}{}_{i,j}^{k}$), and one output which is the weights of features. Fig 6 illustrates an original and selected feature vector.

7- Choose Features, which achieve the Threshold described in equation (9).

$$Threshold = \max(Weight) - \left(\frac{800}{Num}\right) * \left(\max(Weight) - \frac{1}{Num}\sum_{i=1}^{\max(weight)} Weight_i\right) \quad (9)$$

Where Weight is the weight vectors resulting from fuzzy model, 800 is the required number of weights, Num is the number of original features.

2.5.1 Fuzzy Inference System Specifications:

Input1 is the inner difference (Similarity). The membership functions here are 5 triangular functions with values between 0 and 50. While, Input2 is the out difference (Divergence). The membership functions here are 14 triangular functions with values between 0 and 50, and here the functions have opposite effects of the previous functions. The output is the weight score of each feature which will be on a scale from 1 to 15 and each number represents the importance of the feature. We used the triangular functions because the inner and out difference are linear values range from 0 to 50.

2.5.2 Fuzzy Rules:

We configure a lot of fuzzy rules to build the fuzzy system. The rules take in account different situations in which the inner difference and outer difference changes widely from bad to excellent member values. In the following there is an example of the built rules. If (innerDiff is VeryBad) and (meanOutDiff is VeryBad) then (Weight is one).

2.5.3 De-fuzzification:

we applied the “AND” rule to perform the fuzzy operation, and the MAX way to unify the fuzzy outputs. Finally, calculating the output fuzzy of the system is done by “the centroid of area” method.

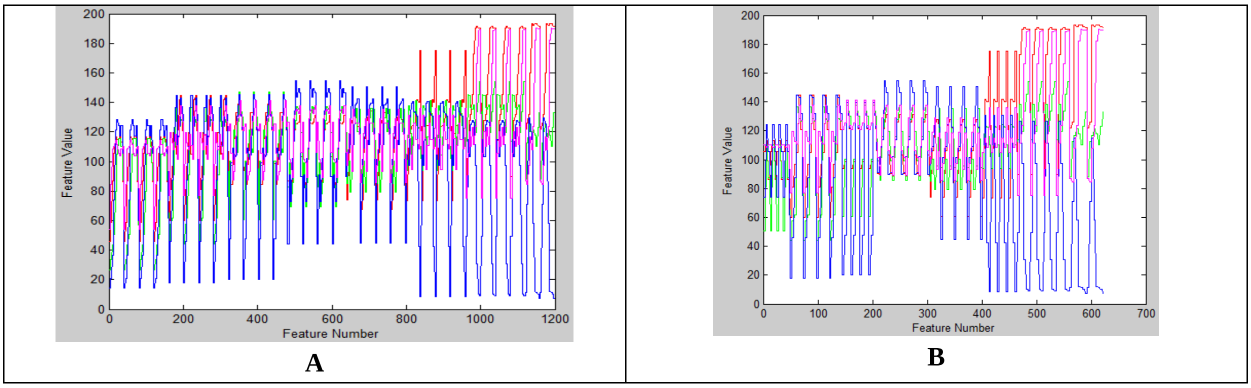

Fig 7. Feature Vector: A. Original, B. Weighted and selected

2.6 Feature Fusion:

The selected features of face and ear are fused together to configure a unified feature vector consists of 16000 face features instead of 60000, and 2000 ear features instead of 8000, and 6900 palm features instead of 15000 . The unified feature vector contains 24900 samples instead of 83000 samples which constitute almost 30% from the original feature vector. This will reduce the required time for the next steps.

2.7 Classification and Recognition:

We suggest using the feed forwarding back propagation neural networks which has the following specifications:

Number of layers:2

Number of neurons in each layer:

- 24900 for input layer which are the number of feature vector samples, 1000 for hidden layer,10 for output.
- Hidden layer function: tansig.
- Output layer function: tansig.
- Training Function: trainlm
- Performance function: Mean Squared Error MSE.

We trained the network using the selected and extracted feature vectors for all individuals in database.

The distances between the NN simulations results of test vector sample and training samples are computed as equation 10 illustrates.

$$d(\vec{x},\vec{y}) = \sqrt{\sum_{i=1}^{N} \frac{(x_i - y_i)^2}{s_i^2}}, \tag{10}$$

The mahalanobis distance of the subtraction indicates to the individual id that the test sample is the closest to and represents the recognized person.

3. Results:

We used ScFace dataset [6] which contains faces and profile images for each individual with different poses (0^0,30^0, 45^0, 60^0,90^0,120^0,135^0, 150^0, 180^0). We selected 1440 images related to 60 individuals (8 images per one) for face, ear and palm training database. We apply tests on face and ear images separately, then we test the fused features. In order to illustrate the effect of nature and size of testing dataset on the performance rate, we decompose the test dataset into 5 different dataset. Table 1 and fig8 illustrate the nature of each dataset.

Table 1. The four different test datasets used in this search.

Dataset	**Size**	**Nature of images**
Dataset01	200 (90 ear, 110 face)	Normal images with different pose variations (30^0, 45^0, 60^0,90^0,120^0,135^0, 150^0, 180^0)
Dataset02	290 (120 ear, 170 face)	Pose Variations, Occlusion with hair and earrings for ear, Partial cutting of face or ear.
Dataset03	616 (216 ear, 400 face)	Adding the frontal images (0^0 pose)
Dataset04	806 (296 ear, 510 face).	Same as Dataset03
Dataset05	1136 (200 ear,468 face, 468 palm print)	Add Palm print to DB 4

Figure 7 includes samples of ear and face images of the four datasets.

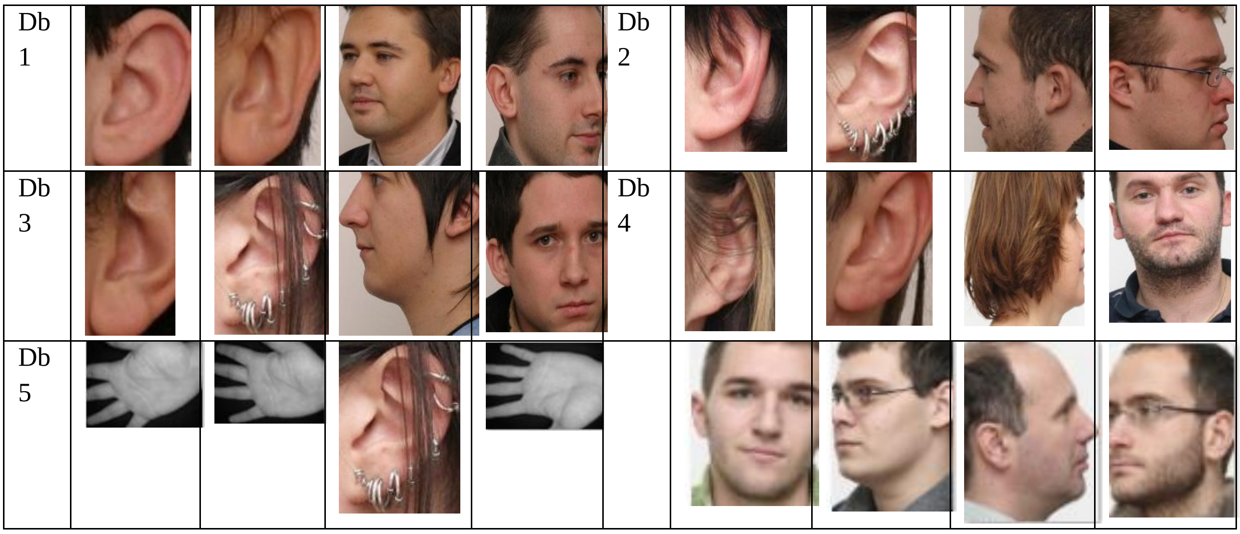

Fig 8. Samples of ear and face images of the four datasets

The recognition rate for ear and face testing datasets separately and fused are shown in table 2.

Table 2. Recognition rate for each dataset.

Type	**Dataset01**	**Dataset02**	**Dataset03**	**Dataset04**	**Database05**
Face	%92.38	%89.3	%89.38	%92.08	%92.08
Ear	%84.5	%80.32	%82.87	%86.04	%86.04
Palm	-	-	-	-	%93.75
(Face +Ear) Fusion	%95.5	%91.46	%93.33	%94.33	-
Fusion (Face + Ear + Palm)	-	-	-	-	%97.4

The recognition rate drops from 92.38% to 89.3% in dataset02 under effect of partial occlusion by hair and earrings in ear images. However, it rises again in dataset03 and dataset04 under effect of adding the frontal face images which are easier to be recognized than other poses. The best result is the fusion result of face, ear and palm, which is 97.4%.

From table 2, we can also see that recognition rate do not degrade under rising up the size of test dataset (i.e. dataset04 is bigger than dataset03 1.3 times but the recognition rate of set04 is bigger than set03) which means that the selected features are independent of datasets size. Also, by adding another biometric, the recognition result is improved.

3.2 Classification Time Comparative:

Table 3 includes comparative results of the classification rate and time at different situations of feature vector for dataset 4.

Table 3 Classification rate and time comparative.

Feature Vector Size	Time (in seconds)	Recognition rate (%)
Hall Feature Vector	0.18	95.3%
First 500 samples	0.042	79.9%
512 selected samples	0.032	82%
16200 selected samples (face+ear)	0.1	94.33%
24900 (face+iris+palm)	0.21	97.4%

Discussion

The result shows the following significances:

- The classification time of the selected features is lower than the hall ones.
- The recognition rate of the selected features decreases between 0.01% to -1.5% in comparing to the original hall features although that the selected features constitute only 26% from the original. In other words, the selected promised features, which constitute only 30% from feature vector, preserved a high performance and decreased the classification time (goal of our research).
- By adding a third biometric (palm), the recognition rate increases by almost 3%.
-

IV. CONCLUSION

The suggested study introduces a new algorithm to select the best features from feature vector of face, ear and palm images in order to preserve a high performance and decrease the classification time. The algorithm can deal with illumination problem by a new technique correcting the illumination by studying the histogram. The segmentation part can take any face image with different poses (frontal or profile) and extract the correct biometric (face or ear), and the classification method depends on training neural network.

The experimental tests show that the selected and weighted features always preserve high performance and low classification time and they are independent off the nature of database, and the system achieved 94.33% and 97.5% recognition rates (face+ear, face+ear+palm) in presence of pose and illumination variations, partial occlusion by hair or earrings and some deforming in test samples.

ACKNOWLEDGEMENTS

Portions of the research in this paper use the SCface Database collected by Capturing face images took place in Video Communications Laboratory at the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia.

REFERENCES

[1] S.M.S. Islam, R. Davies, M. Bennamoun, R.A. Owens and A.S. Mian, *“Multibiometric human recognition using 3D ear and face features”*, Pattern Recognition, Vol. 46, No. 3, 2013, pp: 613-627.

[2] GARCIA E., AGUILAR G., ESCAMILLA E., ***"Face Recognition using Fusion Levels"***, National Polytechnic Institute, ESIME Culhuacan, 2011

[3] Akash P, Avinandan S, Debolina D, Sukanta S, Madhuchhanda D, *“An approach for identification using ear and face biometrics employing score based image fusion and SIFT feature detection in surveillance system”*, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 4, 2015, PP: 74-80.

[4] Songze Lei and Min Qi, *“Multimodal Recognition Method based on Ear and Profile Face Feature Fusion”*, International Journal of

Signal Processing, Image Processing and Pattern Recognition, Vol.9, No.1, 2016, pp.33-42.

[5] R. GONZALEZ. and R.WOODS, *"Digital Image Processing"*, 2nd Edition, Prentice Hall, January 2002, pp: 386-393.

[6] SCFACE Database, Surveillance Cameras Face Database, Available at http://scface.org. Downloaded at: 1-9-2015.

[7] SHAMS Y., TOLBA A., SARHAN S., ***"Face, Iris, and Fingerprint Multimodal Identification System Based On Local Binary Pattern With Variance Histogram and Combined Learning Vector Quantization"***, Journal of Theoretical and Applied Information Technology, Vol.89. No.1, pp: 53-70, 2016.

[8] VAIRAVEL P and VALARMATHY S., ***"Human Identification by Fusing Face, Palm Iris"***, International Science Press, 10(02), 2017, pp.427-435.