Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
COM2012-0049

A Multi-Level Hierarchical Biometric Fusion Model for Medical Applications Security

Sorin Soviany,
Senior research engineer gr. III,
Ph.D. student,
INSCC Bucharest,
Romania

Sorin Puscoci,
Ph.D. Senior research engineer, gr. II, INSCC Bucharest, Romania

Mariana Jurian,
Professor,
Pitesti University, Romania

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209
Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm

Printed in Athens, Greece by the Athens Institute for Education and Research.

ISSN **2241-2891**

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. The papers published in the series have not been refereed and are published as they were submitted by the author. The series serves two purposes. First, we want to disseminate the information as fast as possible. Second, by doing so, the authors can receive comments useful to revise their papers before they are considered for publication in one of ATINER's books, following our standard procedures of a blind review.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Soviany, Sorin, Puscoci, Sorin and Jurian, Mariana (2012) "**A Multi-Level Hierarchical Biometric Fusion Model for Medical Applications Security**" Athens: ATINER'S Conference Paper Series, No: COM2012-0049.

A Multi-Level Hierarchical Biometric Fusion Model for Medical Applications Security

Sorin Soviany,
Senior research engineer gr. III,
Ph.D. student,
INSCC Bucharest,
Romania

Sorin Puscoci,
Ph.D. Senior research engineer, gr. II, INSCC Bucharest, Romania

Mariana Jurian,
Professor,
Pitesti University, Romania

Abstract

The paper proposes a hierarchical biometric data fusion model to provide the performance enhancement of the recognition task in biometric identification and verification systems for medical applications security. The hierarchical approach is relying on more classifiers combination within a multi-level biometric fusion scheme. The multi-level biometric fusion model includes both of pre-classification fusion with optimal feature selection and the post-classification fusion. Our solution increases biometric recognition accuracy based on a suitable feature selection. The novelty of our approach is the combination between feature-level biometric fusion and matching-score/decision-level biometric fusion, ensuring more discriminant information for the authentication task. This approach is suitable for high and medium-security level applications, such as the telemedical ones.

Contact Information of Corresponding author:

1. Introduction

The various applications need custom and optimal solutions to meet various user requirements. Particularly, the security of Internet-based medical applications requires more efficient authentication mechanisms to protect remote medical databases. The actual focus in security systems design and implementation is on biometrics. However, the biometrics do not completely substitute the conventional authentication means, but provide an additional security level. On the other hand, the actual researches are focused on multimodal biometric systems development which are integrating more biometric technologies. This modern approach is intending to reduce the biometric systems inherent errors rates and also to more efficiently handle non-universality of some human traits. The biometrics integration could be done by feature, matching score and even decision-level fusion. Most of actual developed systems were exclusively relying on post-classification and/or matching score-level fusion. [1][2]

The paper proposes a hierarchical biometric data fusion model to provide the accuracy enhancement of recognition task in biometric systems for medical applications security. The model uses a multi-classifier approach for each of the integrated biometrics, embedding feature-level and classification-level biometric fusion. The novelty of our approach is the local and global combination between the two fusion schemes and also the relevant feature selection procedure, giving more discriminative information.

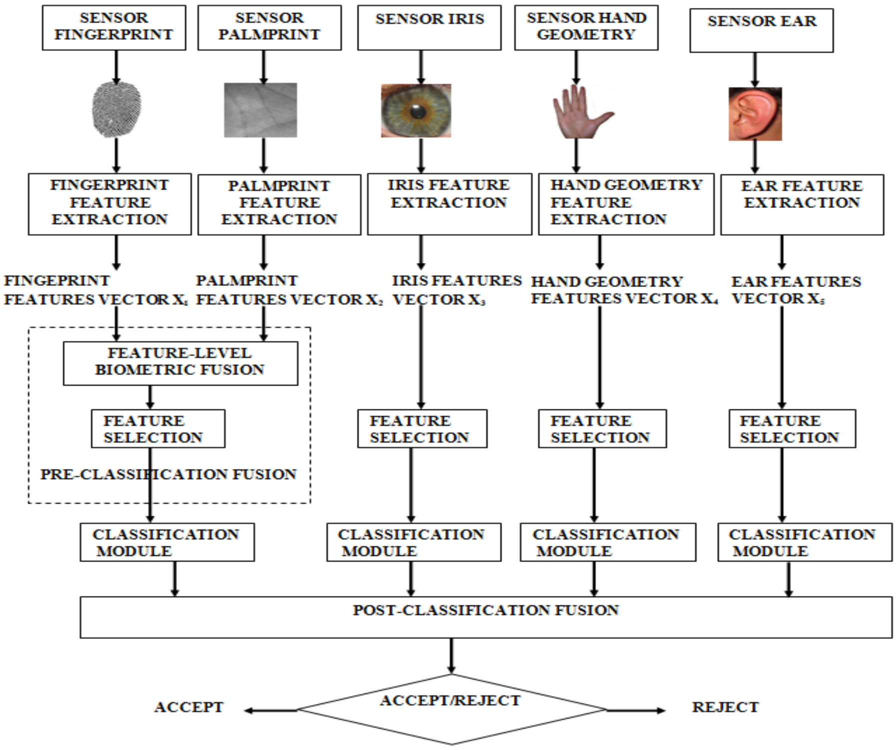
The remainder of this paper is structured as follows. Section 2 briefly presents the proposed system architecture. Section 3 describes the classification models for the integrated biometrics. Section 4 presents our biometric fusion strategy. Section 5 presents the achieved results, based on ROC analysis. Section 6 concludes our research and provides future research directions on biometric systems optimization in order to design more accurate and efficient security solutions.

2.The multimodal system architecture

The security biometric system architecture integrates 5 functional components for people recognition, using the following human traits: fingerprint, palmprint, iris, hand geometry and ear. Figure 1 depicts the general system architecture.

First we will briefly present the ***2 main functional components*** for each of the 5 human traits identification components: *feature extraction with each biometric feature space representation* and *classification component*. Then the ***classification models*** and the ***multi-level fusion strategy*** will be detailed in section 3 and 4 as they are featuring our approach for decision optimization in multimodal biometric systems. We are using a multimodal biometric database with 100 persons human traits records for fingerprint, palmprint, iris, hand geometry and ear.

Figure 1. The multimodal architecture with multi-level biometric fusion and multi-classifier approach



2.1 The feature extraction and feature space representation

For each of the 5 biometrics we applied Principal Component Analysis to extract those features which maintain most of the original data variance. PCA is an *unsupervised* linear feature extractor designed to find a linear subspace from the original data space which retains as much data variance as possible and makes projected data de-correlated. Basically it performs the following tasks: [3][4][6][7]

- computes mean vector μ and covariance matrix Σ for the full data set;
- computes the eigenvectors and eigenvalues;
- chooses the k eigenvectors having the largest eigenvalue;
- obtains the d x k matrix A containing the k eigenvectors;
- projects the initial data into the k-dimensional subspace according to

$$x' = A^T \cdot (x - \mu) \qquad (1)$$

where x is the initial data representation, and x' is the transformed data representation. Also we turned PCA into a supervised feature extractor / dimensionality reduction algorithm by using a *pooled covariance matrix* S_w:

$$S_w = \sum_{i=1}^{C} P_i \cdot \Sigma_i \qquad (2)$$

where:

C is the *classes number*. For biometric verification), C =2, and for identification, C > 2 (each class provides one person identity); P_i is the class i prior (given from the training dataset); Σ_i is the covariance of class i.

For a certain biometric, the ***feature space*** contains all corresponding feature vectors. Table 1 resumes the feature space representation for the 5 biometrics.

Table 1. Feature space representation for the 5 biometrics

Biometric	***Initial dimensionality***	***Final dimensionality***	***Features physical significance***
Fingerprint	25	12	minutiae-related features (ridge ending, bifurcation and dots); distances between 5 relevant points on a central ridge
Palmprint	34	19	distances between the main lines
Iris	41	13	normalized distance between r iris boundaries; spatial location, orientation and frequency for iris patterns, texture details, spots, furrows, stripes
Hand geometry	25	10	geometric parameters for fingers and hand (widths)
Ear	20	11	distances from 3 reference points to the boundaries

The reduced dimensionality values are achieved after PCA application, without any feature selection.

2.2 The classification component

Having the feature space representation for each of the 5 biometrics, our data classification strategy is relying on the following approaches:

- a ***mono-classifier model***: the best classifier selection for each biometric;
- a ***multi-classifier model***: The multi-classifier model is based on 2 different ways:
  - **classifiers fusion:** combination of more classifiers, for the same output type;
  - **classifiers hierarchy**: cascading of more classifiers decisions.

3. *The classification models*

3.1 Basic classification and decision models for system components

First, we applied ***more classifiers*** for each of the integrated biometric (fingerprint, palmprint, iris, hand geometry and ear), in order to find out the *optimal training data set size* and *the best individual classifier*. Their outputs were given as *confidence levels* in class memberships (posterior probabilities or soft outputs) or as *labeling decisions* to provide decision boundaries for each classifier (crisp or decision outputs). Then we performed ***pre- and post-classification fusion*** and we compared the *achieved performances* for the ***multi-classifier approach*** vs. ***the best classifier approach***, with and without the *additional feature selection*. For each biometric, we designed the individual classifiers with the suitable-sized *training datasets* $\boldsymbol{Z_i}$, $= \overline{1,5}$, where the optimal training data set size is resulting from the classifiers learning curves. The new biometric samples are represented by the *feature vectors* $\boldsymbol{x_i}$, $i = \overline{1,5}$, having their dimensionalities according to the values given in table 1.

We considered the following generative classification models: quadratic normal based classifier (***quadratic discriminant***) and ***Parzen classifier***. The discriminant and decision function for these classification models are as following: [3] [5]

- the ***quadratic discriminant classifier (QDC)*** assuming Gaussian density classes, for a 2-class problem, with different class covariances ($\Sigma_A \neq \Sigma_B$). The discriminant function is:

$$R_{QDC}(x) = \frac{1}{2} \cdot (x - \mu_B)^T \cdot {\Sigma_B}^{-1} (x - \mu_B) - \frac{1}{2} \cdot (x - \mu_A)^T \cdot {\Sigma_A}^{-1} \cdot (x - \mu_A) + const \quad (3)$$

where:

$$const = \log \frac{p_A}{p_B} + \frac{1}{2} \cdot \log \frac{\det(\Sigma_B)}{\det(\Sigma_A)} \quad (4)$$

The multi-class extension is performed by averaging the classifier soft outputs for each class pairs within the validation dataset. Alternatively, the multi-class extension could be provided by the *one-versus-all approach*, computing the classifier output for each class, and then giving the final output as the highest normalized score;

- the ***Parzen classifier*** is based on a non-parametric estimation of the class-conditional probability densities, according to the following equation:

$$f_\omega(x) = P(x|\omega) = \frac{1}{n_\omega} \cdot \sum_{z_i \in \omega} K_h\left(\frac{x - z_i}{h}\right) \quad (5)$$

where:
x is the feature vector corresponding to the current tested biometric sample;
z_i is the training feature vector drawn from class i distribution;
ω is the class name (or label);
h is the smoothing parameter for the kernel function K_h;
K_h is the Parzen kernel (or window) function. We used the Laplace kernel, which is completely equivalent to the exponential kernel, except for being less sensitive for changes in σ parameter; also it is a radial basis function:

$$K(x, y) = \exp\left(-\frac{\|x - y\|}{\sigma}\right) \quad (6)$$

(x and y are datapoints in the feature vector space). Therefore, for the 2-class problem, the decision or discriminant function is, in this case:

$$R_P(x) = p_A \cdot f_A(x) - p_B \cdot f_B(x) \quad (7)$$

where p_A and p_B are the classes priors, and $f_A(x)$, $f_B(x)$ are the class-conditional density estimates (based on Parzen window). The **multi-class** extension is performed by providing the following discriminant/decision function:

$$R_p(x) = max_\omega p_\omega \cdot f_\omega(x) \quad (8)$$

We used the following **discriminative classification models**: ***Fisher classifier*** and ***Nearest neighbour rule*** (actually K-Nearest Neighbor classifier). [3][5]

- the ***Fisher classifier***: a linear discriminant which provides the classes separation by finding out a linear transformation w in the data representation space, to maximize the Fisher criterion:

$$J(w) = \frac{\sigma^2_{inter-class}}{\sigma^2_{intra-class}} = \frac{\det(w^T \cdot S_{between} \cdot w)}{\det(w^T \cdot S_{within} \cdot w)} \quad (9)$$

Having the N-sized training dataset Z (for each biometric), the intra-class scatter matrix is given by

$$S_{within} = \sum_{i=1}^{N} \sum_{z \in class(\omega_i)} (z - \overline{z_i}) \cdot (z - \overline{z_i})^T \quad (10)$$

where:

$$\bar{z}_i = \frac{1}{N_i} \cdot \sum_{z \in class(\omega_i)} z \qquad (11)$$

N_i is the number of training samples belonging to class ω_i.
The inter-class scatter matrix is given by

$$S_{between} = \sum_{i=1}^{N} N_i \cdot (\bar{z}_i - \bar{z}) \cdot (\bar{z}_i - \bar{z})^T \qquad (12)$$

where $\bar{z}_i$ is the mean for each class and $\bar{z}$ is total mean vector:

$$\bar{z} = \frac{1}{N} \cdot \sum_{i=1}^{N} N_i \cdot \bar{z}_i \qquad (13)$$

For the biometric identification, the total number of classes is assumed to be the same as the number of training samples. The linear transformation w results from the generalized eigenvalues equation:

$$S_{between} \cdot w = \lambda \cdot S_{within} \cdot w \qquad (14)$$

Finally, the classification is performed in the transformed feature space based on a distance metric; we used Mahalanobis distance because of its main properties (scaling invariance and feature correlation exploiting). Therefore, the new data sample x is classified according to

$$class(x) = argmin_i d_M(x \cdot w, \bar{z}_i \cdot w) \ (15)$$

where $\bar{z}_i$ is the mean of the class i and Mahalanobis distance between 2 data samples x and z (datapoints within the multi-dimensional feature space) is given by

$$d_M(x,z) = \sqrt{(x-z)^T \cdot S^{-1} \cdot (x-z)} \qquad (16)$$

S is the covariance of the 2 data instances.

- the ***KNN classifier***: We applied the KNN classifier in 2 ways. In the 1st approach (*Kappa*), we computed the distance of the new sample to the k-th nearest neighbour per-class, and then assign it based on the minimum obtained distance. In the 2nd approach (*Class-frac*), we estimate the class frequencies among the k-nearest prototypes within the training dataset. We used Mahalanobis distance to find out the closest training neighbours. The choice of K parameter is critical. A higher K value means a smoother, less locally sensitive decision boundary.While K becomes closest to the whole training dataset (N), the classifier performance will approach the most statistical classifiers ones, assigning the class membership to the most frequent class in the training dataset. We minimized the distance influence on the classifier outputs quality by assigning a weight to each neighbour vote; this weight is depending on the distance between the unknown instance and its neighbours within the training set, according to:

$$w(i) = \frac{1}{d_M(x, z_i)^2} \qquad (17)$$

where: w(i) is the neighbour instance z_i weight, and x is the unknown instance.

3.2 The classifiers design for the system components

The classifiers design needs to find out the optimal training dataset size providing the best generalization performance for each biometric sample. This is resulting from the learning curves. The analysis is applied for each of the 5 biometrics and for each of the 4 classifiers (QDC, Parzen, Fisher and KNN). The learning curves for fingerprint, palmprint, iris, hand geometry and ear recognition components are depicted in figures 2,3,4,5 and 6, respectively.

Figure 2. Learning curves for fingerprint classifiers

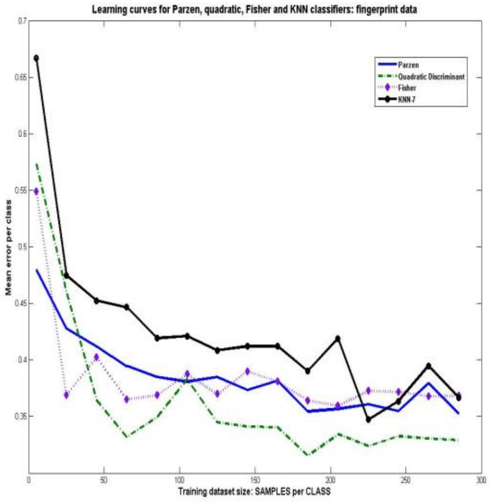


Figure 3. Learning curves for palmprint classifiers

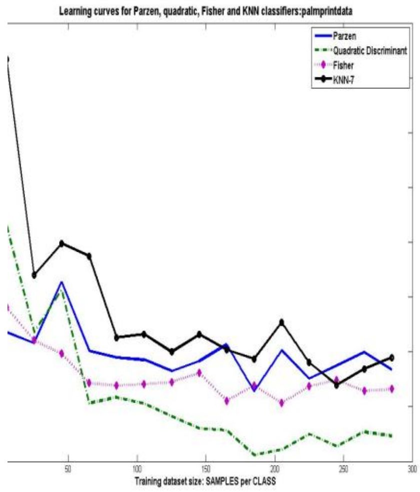


Figure 4. Learning curves for iris classifiers

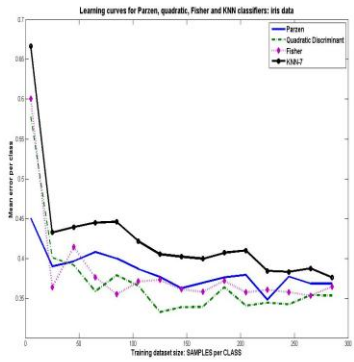


Figure 5. Learning curves for hand geometry classifiers

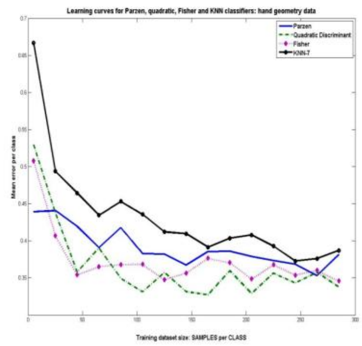
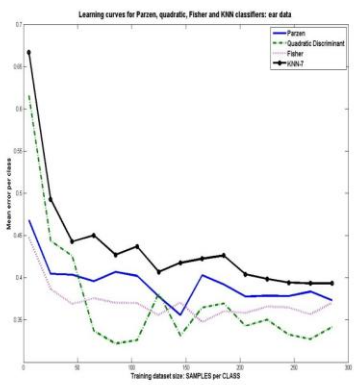


Figure 6. Learning curves for ear classifiers



From these learning curves we could see that the Quadratic discriminant and Fisher classifiers are better performing, especially for larger training sets. However, their generalization error rates are still high. This is the main reason to perform further optimization, either locally and globally. In order to improve the overall system performance, we will apply pre- and post-classification biometric fusion.

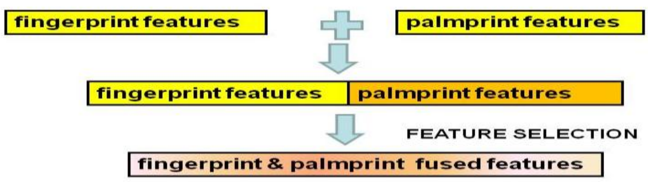
4. The multi-level biometric fusion

We applied a 2-level biometric fusion scheme, which contains: the pre-classification fusion and post-classification fusion.

4.1 Feature-level or pre-classification fusion with feature selection

The pre-classification biometric fusion is performed at feature-level and also it includes the feature selection for the fused biometrics (in our case, fingerprint and palmprint).
Having the feature vectors from fingerprint x_1 and from palmprint x_2, we performed **pre-classification fusion** by their concatenation. This approach is suitable for these 2 biometrics as much as some of their features are commonly originating (minutiae, ridges and so on). Therefore, the basic pre-classification scheme is according to figure 7. This pre-classification fusion scheme includes also the feature selection procedure which is applied on the resulted concatenated features of fingerprint and palmprint. We performed feature selection by applying a *floating search strategy* in which we alternated the *forward* and *backward* selection modes within a fixed number of forward and backward steps. Actually after each forward step we run a backward step and we excluded a feature conditionally if the backward step yields the criterion improvement. The selection criterion was a wrapper one based on a trained classifier error, actually the 1-NN error.

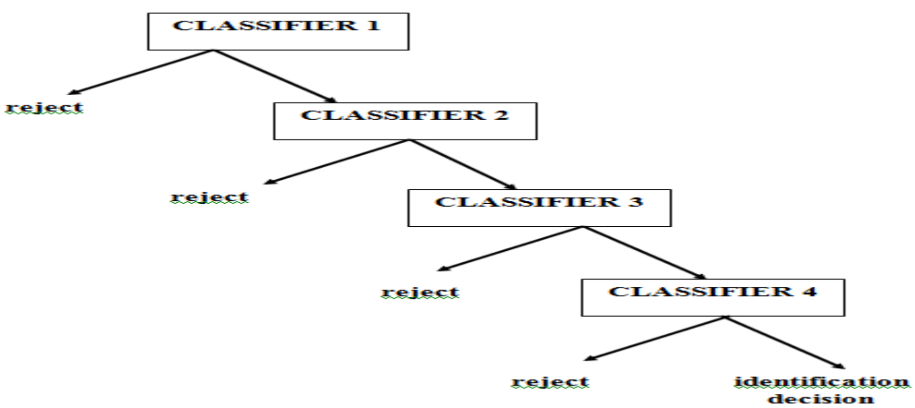
Figure 7. Pre-classification fusion with feature selection



4.2 Post-classification fusion

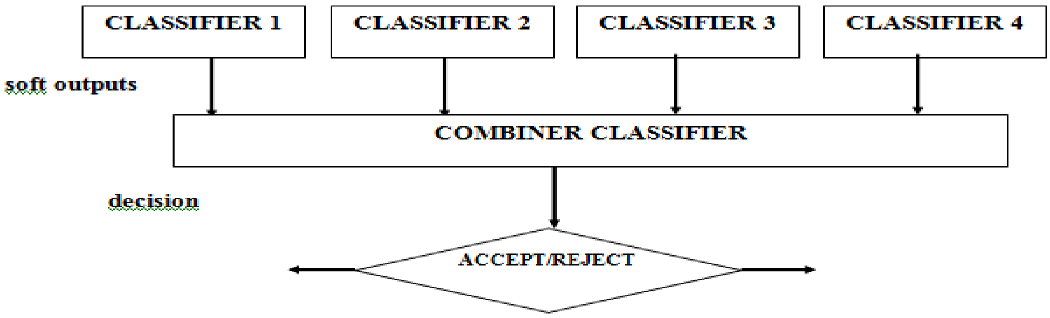
We performed the post-classification biometric fusion in 2 alternative ways, and finally we compared their results.
The first approach is depicted in figure 8 and it relies on a classifiers hierarchy or more decision classifiers cascading. The 4 classifiers considered here are Parzen classifier with Laplace kernel, Quadratic discriminant, Fisher and KNN (with K =7).

Figure 8. Post-classification fusion: classifiers hierarchy decisions cascading



The 2nd approach is shown in figure 9 and it combines the soft outputs of the 4 classifiers, and the final decision is resulting from the overall soft output.

Figure 9. Post-classification fusion: classifiers soft outputs combination



The classifiers are only compared for the same output type; additionally the soft outputs are normalized in order to give the same values range. We applied the normalization based on the *sigmoid* function, because it provides values in range (0,1), which are similarity scores (or posterior probabilities) for the classifiers outputs:

$$Y = p(\omega|x) = \frac{1}{1 + \exp[\, -(A \cdot f(x) + B)]} \quad (18)$$

where f(x) is the classifier output for the input feature vector x, A and B are constants experimentally determined for the actual application.

5. Results, optimization and discussion

The optimization is performed by ROC analysis for fixing the optimal operating point, in order to minimize the classification error rates, especially for the identification task. We used a database containing biometric records from 100 persons. We trained the classifiers with 50 per-class samples, and we perfomed the validation on an independent data set containing 20 per-class samples. Figure 10 presents the ROC curve with error rates on 2 persons with their biometric data being drawn from our database: person A and person B. The ROC analysis is performed on post-classification fusion with hierarchical approach, without any additional feature selection. Figure 11 presents also the ROC analysis for the same hierarchical approach on post-classification biometric fusion, but with feature selection.

Figure 10 : ROC curve for post-classification fusion with hierarchical classification without feature selection

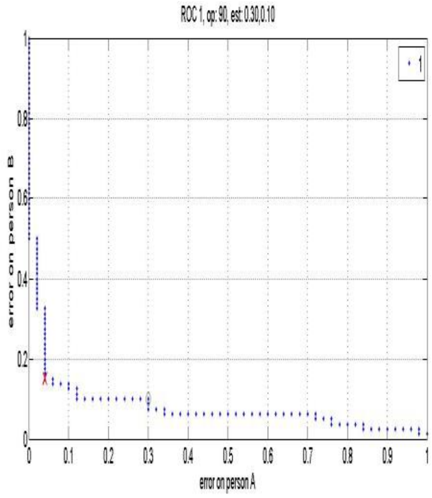
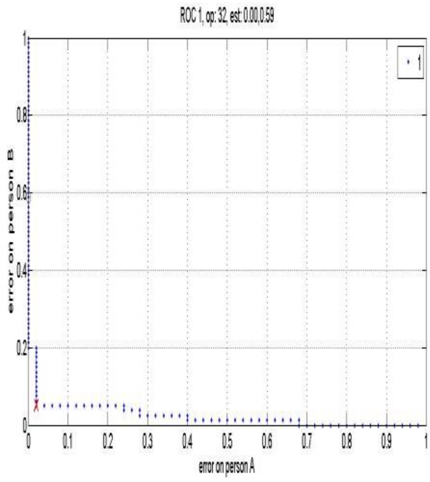


Figure 11 : ROC curve for post-classification fusion with hierarchical classification including feature selection



In the 1st case, the average error achieved for 2 persons identification on optimal operation point was 0,095. In the 2nd case, the optimal selected operation point provides an average error of 0,035.

Figures 12 and 13 depict the same analysis performed on post-classification fusion, without and with feature selection, respectively, but considering the 2nd fusion strategy based on classifiers soft outputs combination (with the weighted sum rule).

Figure 12 : ROC curve for post-classification fusion with outputs combination without feature selection

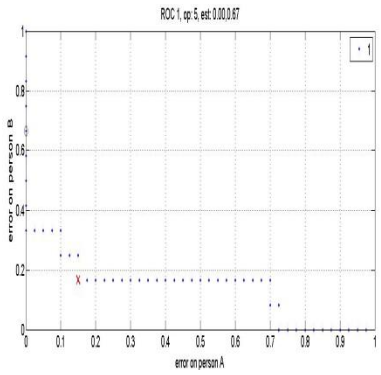
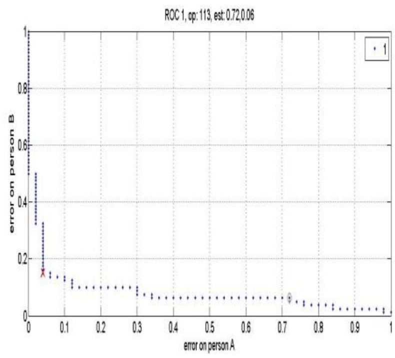


Figure 13 : ROC curve for post-classification fusion with outputs combination including feature selection



By applying the classifiers combination, the per-class average error rate was 0,16 without any feature selection (according to figure 12), respectively 0,09 while we applied the additional feature selection procedure.

In both post-classification fusion, the additional feature selection procedure improved the persons identification performance.

6. Conclusions and further research

The recognition task could be efficiently performed with various complexity classifiers, depending on the available data set size. The low-complexity classifiers provide better generalization performances than more complex ones for small-sized training datasets. On the other hand, classifier optimization need often to keep some trade-offs as much as there are not an ideal classifier. However, the performance improvement achieved by combining more biometric data classifiers provides more security in many authentication-based applications, including applications which are designed for telemedical databases remote access control.

We proposed a multi-level biometric fusion model including not only the post-classification but even the feature-level or pre-classification biometric fusion. This is one of out approach novelty. Additionaly, we shown that a careful feature selection is able to more improve the biometric identification accuracy, as much as not all the feature vectors component provides the same precision. The primary biometric data are often dealing with noise and other conditions, which are increasing the error rates.

Further research has to more carefully approach the feature-level biometric fusion, because of its performance improvement potential.

7. References

[1] Soviany, S., Soviany, C., Jurian, M., ‘A Multimodal Approach for Biometric Authentication with Multiple Classifiers’, International Conference on Communication, Information and Network Securiy, November 28-30, 2011, Venice, Italy

[2] Soviany, S., Jurian, M., ‘Multimodal Biometric Securing Methods for Informatic Systems’, 34th International Spring Seminar on Electronic Technology, May 12-14, 2011, Slovakia

[3] Theodoridis S., Koutroumbas, K. (2009) *’Pattern Recognition’*, 4th edition, Elsevier Inc.

[4] Shlens, J. `Tutorial on Principal Component Analysis`, Center for Neural Science, New York University, 2009

[5] Polikar,. R. `Pattern Recognition’, Wiley Encyclopedia of BioMedical Engineering, 2006

[6] Jain, A.K. `An Introduction to Biometric Recognition`, IEEE Transaction on Circuits and Systems for Video Technology, Special Issue on Image- and Video-based Biometricsm, vol. 14, no. 1, 2004

[7] Snelik R., Indovina M.,`Multimodal biometrics: issues in design and testing’, Proceedings of 5th International Conference on Multimodal Interface, 2003