The Multimodal Fusion in Security Systems Design

Sorin SOVIANY[1], Sorin PUSCOCI[1]

***Rezumat**: Articolul abordează problematica fuziunii multimodale ca tehnică de proiectare şi dezvoltare a sistemelor de securitate. Se realizează şi un studiu de caz pentru un model de fuziune multi-nivel, atât la nivel de caracteristici cât şi la nivel de scor, în cazul unei aplicaţii biometrice. Modelul propus de fuziune multi-nivel asigură o creştere a performanţelor în ambele moduri operaţionale, verificare şi identificare. Componenta de fuziune la nivel de caracteristici are contribuţia cea mai importantă la creşterea performanţelor.*

***Cuvinte cheie**: fuziune multimodală, caracteristici, sistem de securitate, reguli de combinare*

***Abstract**: The paper approaches the multimodal fusion as a design and development technique for security systems. Also there is a case study for a multi-level fusion model, both on feature and matching score level, with a biometric application. The proposed multi-level fusion model provides a performance enhancement in both operational modes, verification and identification. The feature-level fusion component provides the most important support for the performances improvement.*

***Keywords**: multimodal fusion, features, security systems, combination rules*

1. INTRODUCTION

The security systems design and development deal with a lot of challenges, especially while processing information from several sources. This is particularly true for multi-biometric systems that integrate several biometric traits. This integration is done at several processing stages, either pre- or post-classification [1], [2]. Actually there are various rules to combine data from multiple sources in order to meet the real applications requirements. These rules are based on mathematical models that are applied with or without some parameters depending on the real operational conditions.

The suitable fusion rule selection remains a complex decision making issue, and this is due to the various end-users applications technical and cost-related requirements

The remainder of this paper has the following structure: *Section 2* presents the typical fusion rules; *Section 3* presents a case study for a biometric application; *Section 4* concludes about the fusion rules for security systems design.

2. THE TYPICAL MULTIMODAL FUSION RULES

The **multimodal fusion** is the process in which several data samples are combined using some mathematical models, either parametric or non-parametric [1], [3]. These rules are named **multi-modal fusion rules**. The result of a multimodal fusion is an overal matching score or a complete feature set, depending on the processing level in which the fusion rule is applied.

The multimodal fusion concept was applied for biometric systems [1], [3], [4], [5], [6]. However, the multimodal fusion rules should be used for any other

[1] Institutul Naţional de Studii şi Cercetări pentru Comunicaţii, Bucureşti, România.

application requiring pattern recognition/classification in order to meet the desired goals, such as early recognition of some security events occcurece within a local network.. A more general concept is that of **multimodal analysis** including *the multimodal fusion rule* together with a certain *interpretation of the data*. The main applications include the following:

● the **multimodal analysis** for *real-time recognition of network intrusions*, case of *IDS/IPS (Intrusion Detection/Prevention Systems)*. The multimodal fusion rules apply on data featuring various security patterns or malware attacks signatures;

● the **multimodal analysis** for *persons recognition based on their physical and/or behavioural traits*, case of *biometric systems*. The multimodal fusion rules inputs are either the extracted features, matching scores or values of the implemented discriminant functions, eventually decisions (but with a significant information loss cost);

● the **multimodal analysis** for *human activities or even specific human behavioural models recognition*, case of various *smart environments applications*, applications from personal and/or public safety with significant social and/or economical impact. The multimodal fusion rules apply on dynamic feature sets or patterns, with temporal variability and some relationships among their belonging classes.

The typical multimodal fusion rules are divided into the following main classes [2], [3], [6]:

● ***the low-level*** or ***pre-classification fusion***: the task performs during an early stage of data processing. The fusion rule is applied before any advanced processing task that should recognize the similarity or matching elements against the references data. The ***pre-classification fusion*** schemes are the following: *sensor-level* (or *sample-level) fusion* and *feature-level fusion*. The main advantage of the ***low-level fusion*** is that it exploits the highest amount of information from the primary sensor data. This should provide a high accuracy of the patterns classification. However, there is a significant risk to generate low-quality patterns because of the noisy data and significant outliers of the security features values;

● ***the high-level*** or ***post-classification fusion***: operates at a high-level of data processing, just after the matching scores evaluation, before or after decision making. The typical ***post-classification fusion*** schemes are the following: *classifiers fusion; matching score-level fusion* and *decision-level fusion*. The ***high-level fusion*** is the most adopted for real applications because of its easy implementation. However, because of the more advanced stage of data processing, the task exploits less informative data.

2.1. Pre-classification Fusion Rules

2.1.1. Sensor (or sample)-level Fusion

The general structure of a multimodal system with sensor (sample)-level fusion is depicted in figure 1. The primary date from various sources are combined into a single sample. It is not mandatory to process different data types [2], [3]. A face recognition system could combine facial images from several cameras [3], [6].

O sample-level fusion rule is reliable only if the primary data samples are compatible. This is not always true, actually it is an exceptional case for many applications. For instance, the combination of facial images from different resolutions camera is not feasible.

2.1.2 Feature-level Fusion

The feature-level fusion combines several feature vectors provided either from different sensor data or by applying various feature extraction algorithms[3], [5]. The general model of feature-level fusion is depicted in figure 2.

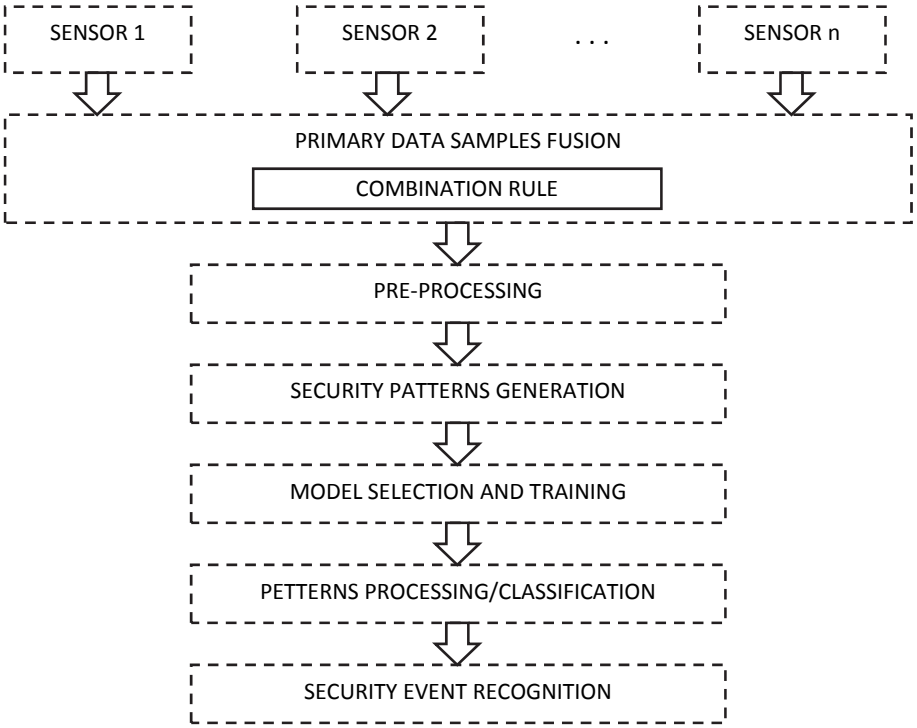


Fig. 1. Multimodal security system with sample (or sensor)-level fusion.

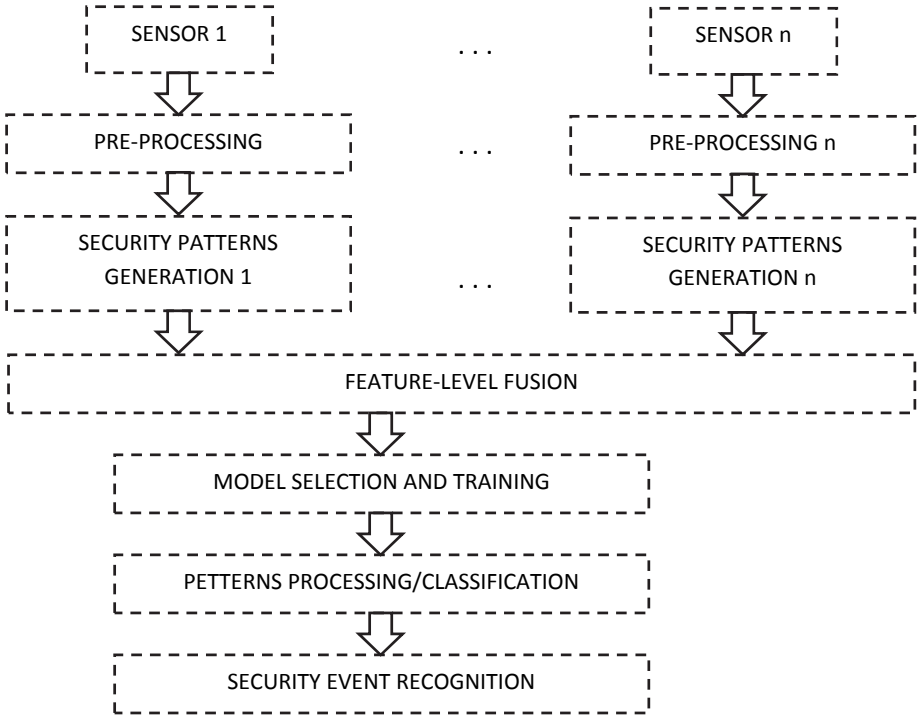


Figure 2. Multimodal security system with feature-level fusion.

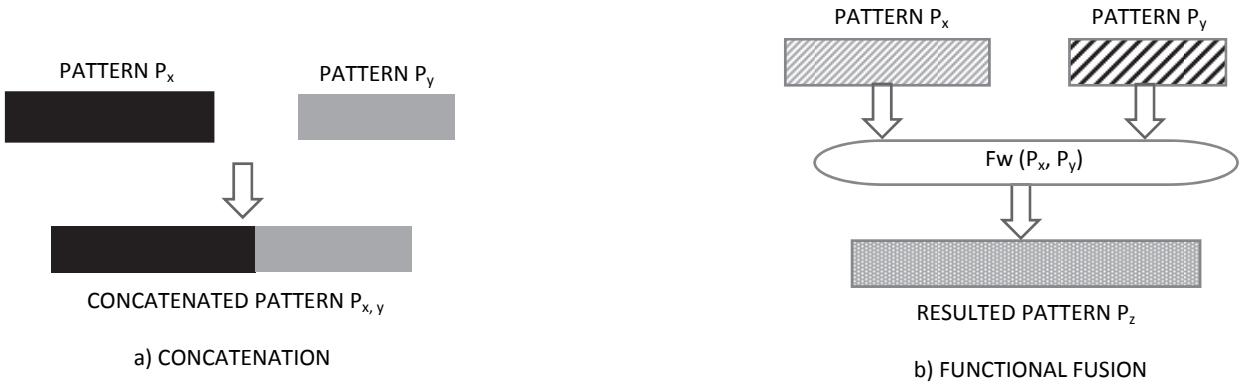


Fig. 3. Feature-level fusion implementation methods.

The fusion of several feature sets generates one pattern. The feature-level fusion provides for the further classification stage an integrated feature set containing several independent features and exploiting more relevant information.

The feature-level fusion could be implemented in the following ways [3], [6], [7], as shown in figure 3:

● **feature vectors concatenation** (figure 3a): ***The concatenation-based fusion rule*** is applied when the feature vectors are *not homogeneous*. The features *heterogeneousness* typically result from the different pre-processing methods for feature extraction or generation task. This approach allows for different patterns sizes. If the original feature sizes are big, the resulting dimensionality will increase exceeding an acceptable threshold for a given computational complexity. This dimensionality expense is the major drawback of the ***low-level multimodal fusion with feature concatenation***. An important requirement for the results consistency is that the feature values for different data should be quantified on the same scale. On the other hand, the concatenation-based fusion is not feasible if the features are incompatible. The security features *incompatibility* usually concerns their physical significances. For instance, in a multimodal biometric system integrating fingerprint and face recognition, the minutiae-based features (that are extracted from fingerprint) and eigenface coefficients (for the face image) are truly *incompatible* [8];

● **feature vectors functional combination** (figure 3b): ***The functional combination of feature vectors*** is a fusion rule that requires the feature vectors *homogeneity*, also with the same sizes. The homogeneous feature vectors results from multiple samples acquisition with the same data type, with the same sensor data pre-processing [3], [7]. ***The functional combination of feature vectors*** involves a mathematic rule applied to each vector component. For the 2 patterns P_x, P_y (in figure 3b), represented by 2 feature vectors (with d their common dimensionality), the functional combination-based fusion generates the resulting pattern P_z according to

$$P_z = Fw(P_x, P_y) \tag{1}$$

where Fw is the combination function, sometimes with a parametric form. The most common functional fusion is the weighted averaging of the patterns components, with the general model shown in figure 4.

The feature-level fusion implementation is constrained by the following issues [3], [6]:

● *uncertainty in finding relationships among the various feature spaces* that should be integrated. It is difficult to find out reliable correlations among the

various extracted features. These functional relationships among the various features could allow a suitable adjustement of the overall feature space, with advantages concerning processing time [9];

● *the increasing of the dimensionality feature spaces*, especially for the concatenation-based fusion. This increases the computational complexity for data processing in classification stages.

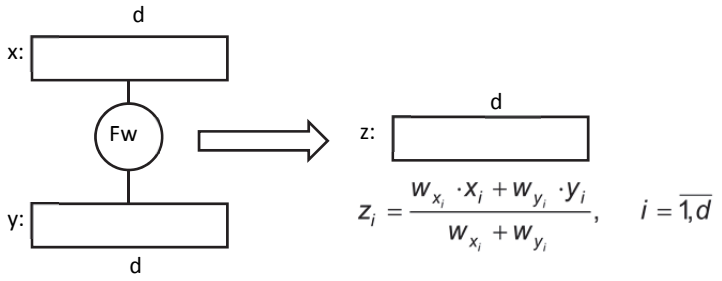


Fig. 4. The weighted averaging-based feature-level fusion [2].

2.2. Post-classification Fusion Rules

The general architecture of the multimodal security systems with high-level (post-classification) fusion is depicted in figure 5.

The key difference between the high-level and low-level fusion is due to the processing stage in which the fusion is applied. A combination process that performs before the matching scores computation exploits more information related to the primary sensor data, even before or after feature extraction. However, the low-level fusion performance is often decreased by the noisy data, especially when the fusion is applied just before any other pre-processing operation.

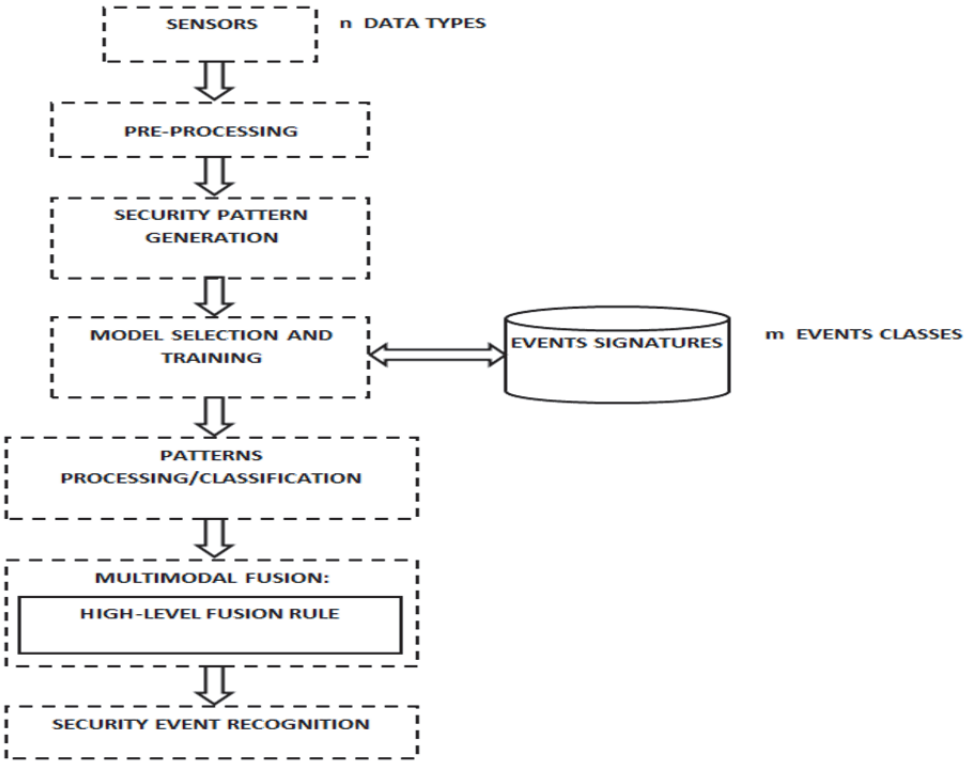


Fig. 5. Multimodal security system with high-level fusion.

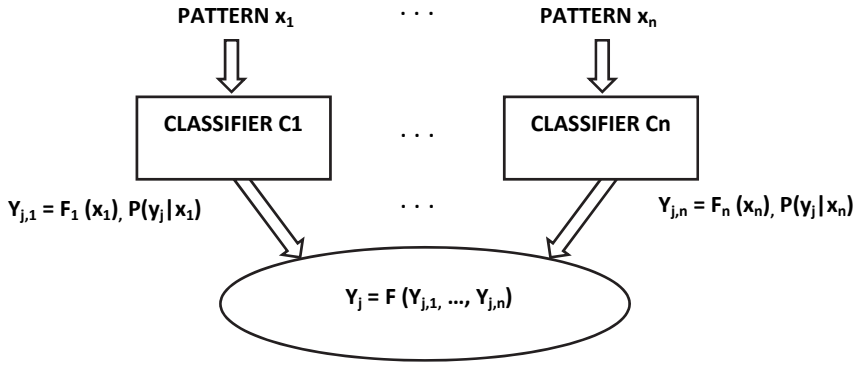


Fig. 6. The general model of the multi-classifier fusion for multimodal security systems [2].

This is the main reson for which the high-level fusion is the most implemented in the actual multimodal security systems. The post-classification fusion is typically performed either by combining the matching scores or the individual decisions. The matching score combination is a mathematical one in which the several individual scores are input data for a certain function. The decisions fusion is applied with some voting schemes, either weighted on unweighted.

2.2.1. Matching Score-level Fusion

The **matching score-level fusion** is the process of several scores combination leading to an overall score that supports the final decision. The scores combination is performed using a certain mathematical model, with or without some parameters.

The matching scores are computed in the following ways:

● **output values of classifiers discriminant functions**, sometimes with an additional normalization. This approach is called **multi-classifier** (figure 6) and it is based on supervised learning. Each of the n discriminant functions computes a score that evaluates the class membership for the current feature vector. The amounts $Y_{i,\ j}$ are the return values of the discriminant functions, where j is the class index, $j = \overline{1,C}$ and i is the classifier index, $i = \overline{1,n}$. C is the overall number of security events to be recognized (or the number of enrolled users to be identified). Each classifier output is a probabilistic score:

$$Y_{j,i} = F_i(x_i) = f(P(Y_j \mid x_i)) \tag{2}$$

This is the posterior probability for which the feature vector x_i belongs to class Y_j. Each class could be a security event or an enrolled identity, depending on the application. The multi-classifier fusion rule outcomes are provided according to:

$$Y_j = F(Y_{j,1},...,Y_{j,n}) \tag{3}$$

where Y_j is the result of the individual classifiers outputs combination with the rule F. For the **multi-classifier approach**, the most common used ***combination rules*** are the following ones [1], [2], [3], [7]: *the class posterior probabilities product*, *sum*, *maximum* and *minimum* rules, respectively (with their normalized forms), *the class posterior probabilities average* and *weighted average* rule, respectively;

● **normalized distances within the multi-dimensional feature spaces**, with a direct comparison between the current feature vector and the reference one (figure 7). This is the true **matching process**, usually more accurate than the supervised learning, but with a significant additional complexity cost, especially for large-scale biometric identification systems.

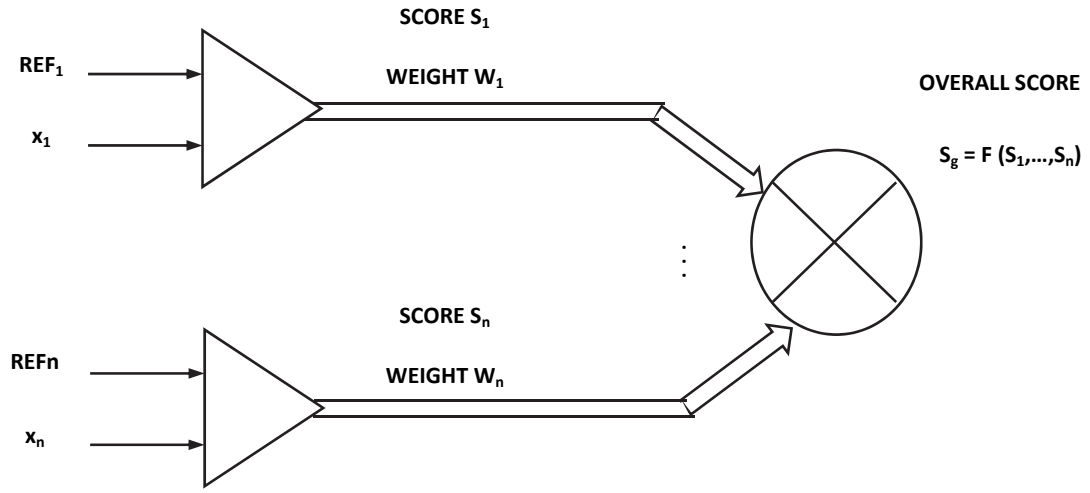


Fig. 7. The general model of the score-level fusion for multimodal security systems [2]

The actual methods for global score computation based on the set of individual matching scores are grouped into the 2 following classes:

● ***parametric models***: The **parametric models** for the score-level fusion are **mathematical rules** in which the ***overall matching score*** S_g is the result of some functions evaluation, having as inputs *the individual scores* and together with *a set of parameters representing weights* that are associated to the individual scores. The general form of the scor-level parametric fusion rule is [2]:

$$S_g = f(S_1,...,S_n;w_1,...w_n) \quad (4)$$

in which:

S_g is the global matching score for all the integrated security patterns types;

$S_i,\ i=\overline{1,n}$ is the matching score that is achieved for the pattern type *i*;

$w_i,\ i=\overline{1,n}$ is the weights for the individual score S_i;

f is the functional combination rule that is choosen depending on the application.

The most common parametric models for the overall matching score computing are the following ones [1], [2], [3], [7]: *the weighted sum*, *product* and *average* rules, respectively.

● ***non-parametric models*** for matching score-level fusion: The **non-parametric models** for the score-level fusion are **mathematical rules** in which the ***overall matching score*** S_g is the result of some functions evaluation, having as inputs *the individual scores*, but without any additional parameters or weighting. The general form of the scor-level non-parametric fusion rule is [2]:

$$S_g = f(S_1,...,S_n) \quad (5)$$

The typical non-parametric models for the non-parametric score-level fusion are the following one [1], [2], [3], [7]: *the sum*, *product*, *average*, *minimum score* and *maximum score* rules, respectively.

2.2.2. Decision-level Fusion

In the **decision-level fusion**, each classifier/ matching module makes an individual decision concerning the security event to be recognized (or person identity to be identified). The overall final

decision is provided by one of the folowing common applied fusion rules [2]: *voting* or *weighted voting*, *AND*, *OR*, *EXCLUSIVE OR* rule, respectively.

The decision-level fusion exploits the lowest amount of information while comparing with all the other fusion schemes (score-level fusion, feature-level fusion) [1], [3], [8]. Therefore, the expected accuracy in event or persons recognition shouuld be lower than for the other fusion rule classes.

3. A CASE STUDY FOR BIOMETRIC APPLICATIONS

Let's consider a multimodal biometric system that integrates 3 biometrics traits: fingerprint, palmprint and iris, respectively [10].

This case study presents:

- the overall system architecture;
- the multimodal fusion model;
- the performance achievements.

3.1. The System Architecture

Figure 8 depicts the overall architecture of the multimodal system for this example. The authentication process require the following system functions [10]:

- ***the feature generation*** for each biometric, with *feature extraction*, *local feature-level fusion* rule and *feature selection* for dimensionality reduction;
- ***the hierarchical classification*** for each biometric;
- ***the post-classification fusion rule*** which makes the decision for the application.

3.2 The Multimodal Fusion Model

For this system we considered the pre-classification and post-classification fusion, respectively.

The ***pre-classification*** fusion is applied within the **feature generation** stage and the ***post-classification*** fusion is related to the biometric data **hierarchical classification** operations.

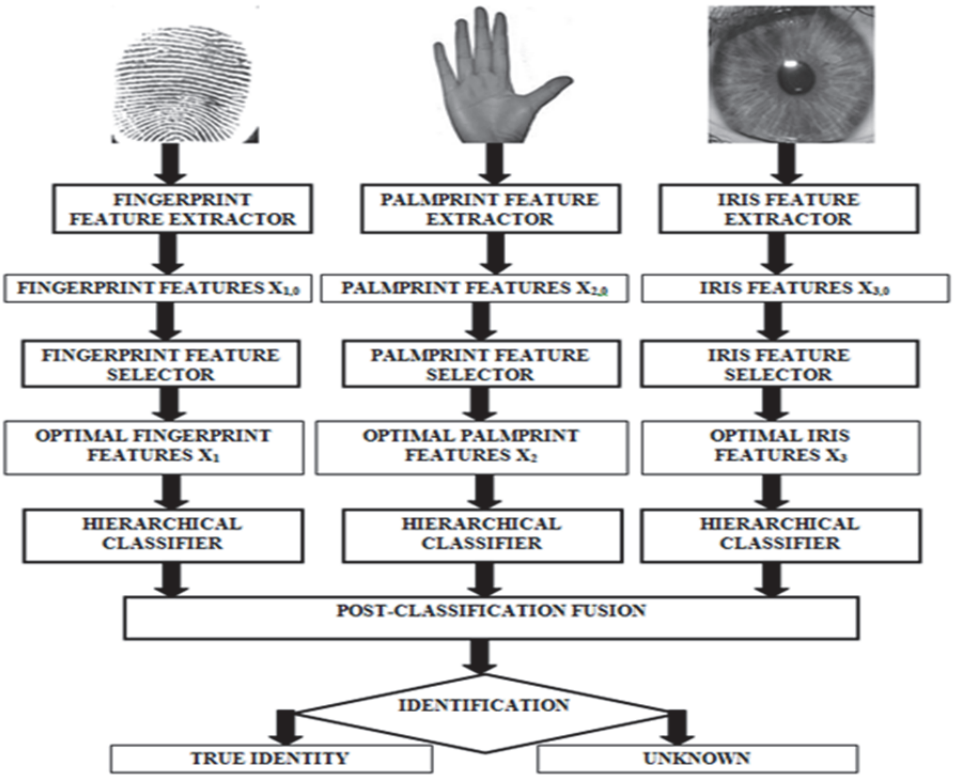


Fig. 8. The Multimodal Biometric System Architecture [10].

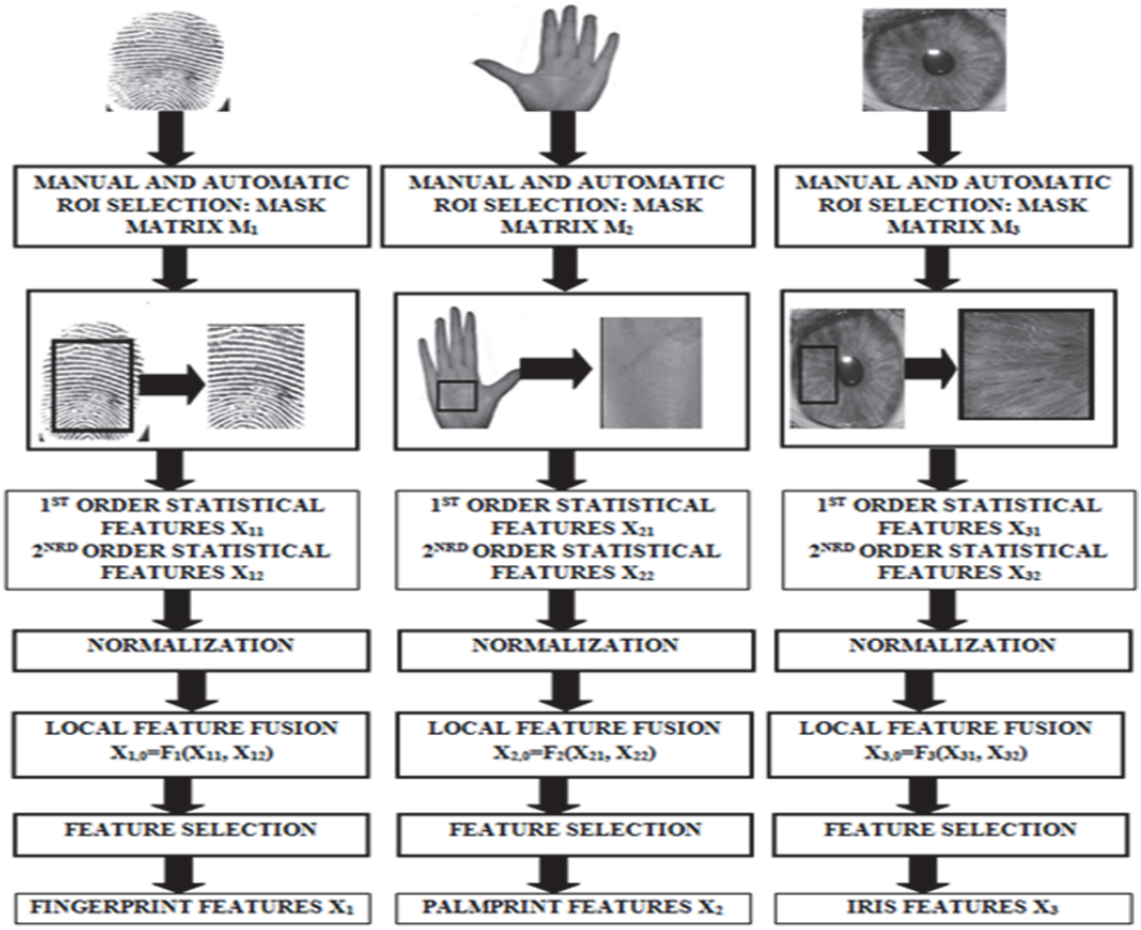


Fig. 9. The Feature Generation [10].

The **feature generation** process is briefly depicted in figure 9 [10].

The **feature generation** stage include a ***local fusion rule*** for the feature subsets that are generated for each of the 3 biometrics. Therefore this is a ***pre-classification*** or ***feature-level fusion***. The feature generation stage has the following steps, according to figure 9 [10]:

- *Manual and automatic selection for the Regions of Interest (ROIs)*;
- *The 1st and 2nd order statistical features computation from the image ROIs pixels*;
- *Discarding of the less correlated features*;
- *Feature normalization*;
- *Local Feature-level Fusion* for each biometric;
- *Feature selection for further dimensionality reduction.*

The local feature-level fusion task is separately performed for each of the 3 biometrics: fingerprint, palmprint and iris. After the feature extraction process we get 2 feature subvectors for each biometric ($k = 1$ for fingerprint template, $k = 2$ for palmprint template and $k = 3$ for iris template): $X_{k,1}$: the 1st order statistical feature subvector; $X_{k,2}$: the 2nd order statistical feature subvector; $X^{*}_{k,2}$: the 2nd order statistical feature subvector after less correlated features discarding.

The local feature-level fusion process is depicted in figure 10. This is a **concatenation** process of the normalized features for each of the 3 biometrics.

The resulting feature sizes are given in table 1.

Table 1

Feature spaces dimensionality [10]

Biometric	k=1: Fingerprint	k=2: Palmprint	k=3: Iris
Size(X_{k1})	9	9	9
Size(X_{k2})	20	29	40
Size(X^*_{k2})	9	10	13
Size($X_{k,0}$)	18	19	22
Size(X_k)	**8**	**9**	**10**

We applied also a ***post-classification fusion rule***; this is a ***decision-level fusion*** based on **weighted voting**. The classification stage is depicted in figure 11 [10].

We designed and trained a hierarchical classifier for each biometric according to the differentiated security requirements for individual recognition: a detection classificatin followed by a discrimination among the other enrolled identities.

In the fusion rule, each individual biometric has its own weight in the final decision. The weights are updated according to the following rule [10]:

$$w_k(i+1) \leftarrow w_k(i) \cdot \alpha_k(i), k = \overline{1,3} \qquad (6)$$

where:

$w_k(i)$ represents the weight that is assigned to biometric pattern k for the iteration step i;

$$\alpha_k(i) = \frac{TPR\det_k(i)}{TPRdisc_k(i)} \qquad (7)$$

$\alpha_k(i)$ is the performance ratio detector vs. discriminant for the iteration i. The applied performance measure is True Positive Rate (TPR). The iteration stops when a certain performance threshold is reached, according to the real security application requirements.

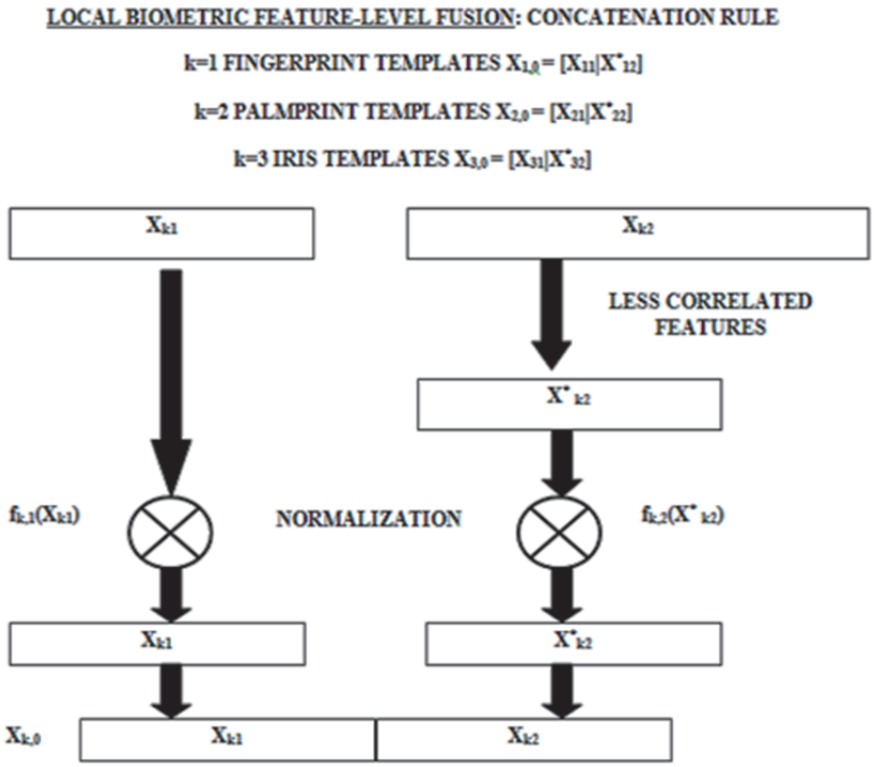


Fig. 10. Local Feature-level Fusion with concatenation [10].

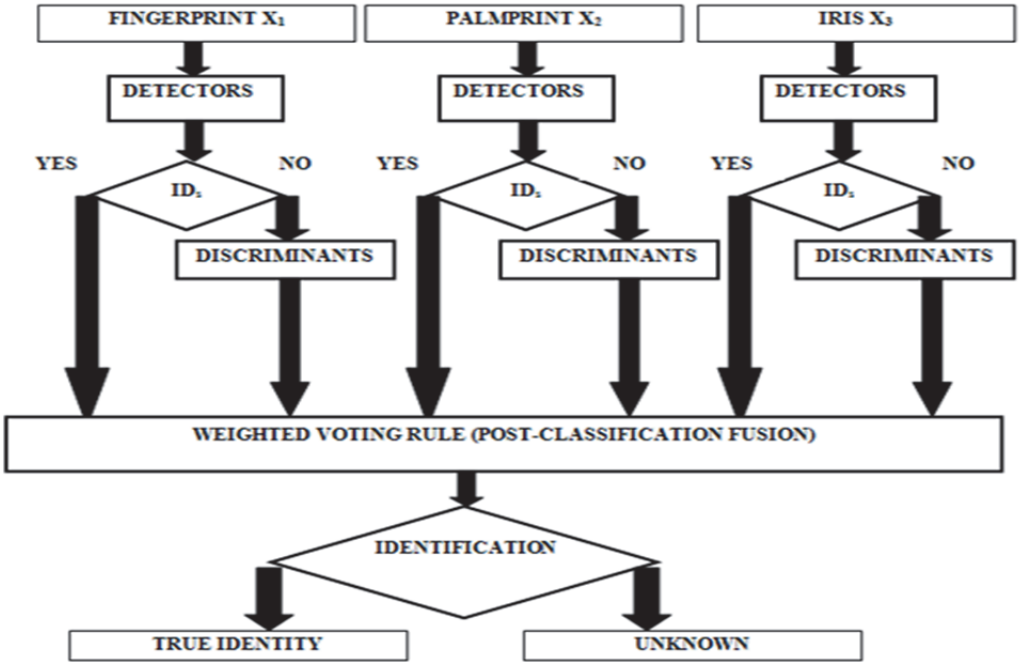


Fig. 11. The classification stage [10].

3.3. The Performance Achievements

For the detectors we applied Gaussian Mixture Models with separate training according to their target identity. For the discrimination stage we considered the following 3 classification models: ***Naïve-Bayes***, ***Parzen*** (with Laplace kernel) and ***Quadratic***. Based on their learning curves for the available data, we selected the suitable models for each of the 3 biometrics and for a training set having only 25 samples per class, as following: for fingerprint Parzen, for palmprint Quadratic Discriminant Classifier and for iris Naïve-Bayes classifier [10].

The system performance is evaluated for 10 experiments with their results averaging. The overall system performance is optimized using ROC (Receiver Operating Characteristic) curves and selecting its optimal operating point according to the application requirements. We considered the identification error rate for some target identities or users which the highest authorization degree. The achievements are the optimal operating points for the overall multimodal system with concatenation fusion rule and with/without detection stage, according to figure 12 and table 2 [10].

Table 2

Optimal operating points for the multimodal biometric system with concatenation feature-level fusion [10]

Op. Point	Identif. Error Rate on person I_1	Identif. Error Rate on person I_2	Avg. Identif. Error Rate
A0	**0.0875**	**0.065**	**0.07625**
B0	0.160	0.09375	0.126875

The designed system performance is significantly improved with a hierarchical classification approach but also applying both multimodal fusion rules, the pre-claassification post-classification fusion, respectively.

The first one is performed at feature-level by concatenating the 2 feature subvectors for each of the integrated biometrics. The overall system performance is significantly higher than for each of the 3 biometrics. The local feature-level fusion supports this performance improvement but only for

the transformed feature sets. On the other hand, the hierarchical classification of biometric data together with the decision-level fusion provides also a certain degree of performance improvement. This performance improvement is achieved using an iterative procedure in which the weights are updated for the weighted voting scheme. Further improvements should be done depending on some particular applications requirements, especially while considering certain security thresholds.

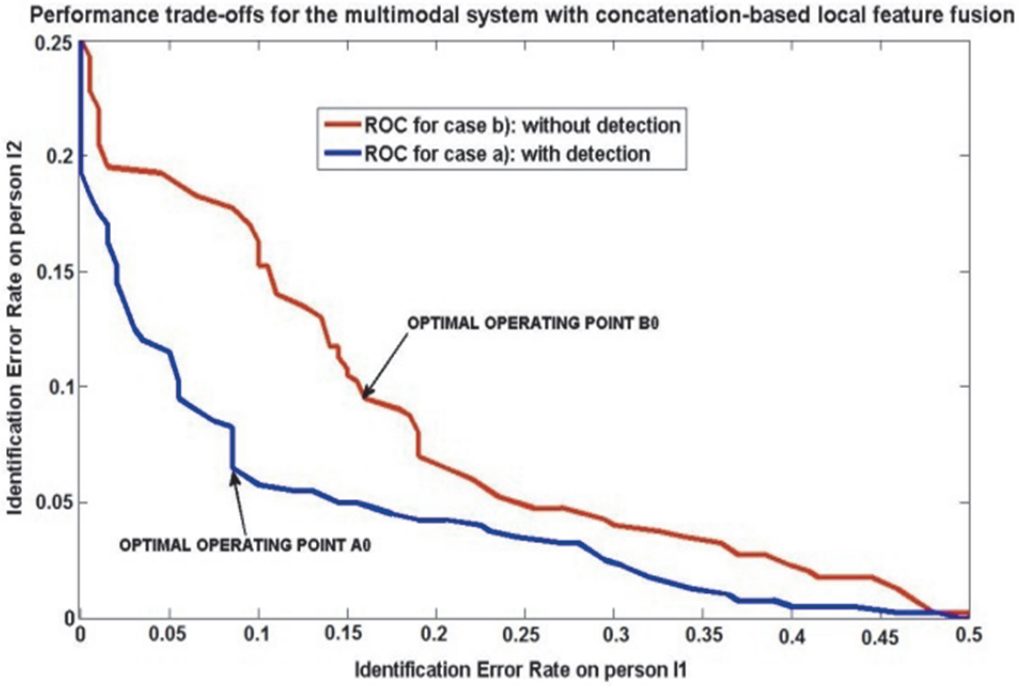


Fig. 12. ROC curves and optimal operating points for the multimodal system with concatenation feature fusion [10].

4. CONCLUSIONS

The actual security systems design and development oftern requires various and sometimes complexity expensive information processing tasks in order to make suitable decisions according to their applications requirements. The information should be combined and analysed in order to define and provide some common security patterns for the further classification process.

Despite of the actual technological advances, there is no universal rule that should be applied for any systems and for any end-users requirements. The typical multimodal fusion rules should only be applied for certain application-related requirements, with their specific security thresholding and also with some parametric models; the models parameters are typically fixed or selected based on the available experimental data.

On the other hand, the various multimodal fusion rules performance achivements are strongly dependent on the data processing stage in which they are applied. This is also supported in our case study for a multi-biometric system in which we applied either a local feature-level fusion for each of the integrated biometric data and a decision-level fusion. This mixed fusion process provided a significant performance improvement for the identification task, but with some particular conditions related by the feature space dimensionality and the training set size.

Further developments are expected to be provided expecially for feature-level fusion, because it can exploit a lot of relevant information that are derived from the extracted features. This is especially the case of biometric applications.

REFERENCES

[1] D. Zhang, F. Song, Y. Xu, Z. Liang: *Advanced Pattern Recognition Technologies with Applications to Biometrics*, Medical Information Science Reference, IGI Global, 2009.

[2] S. Soviany: Ph.D. Thesis *Decision Optimization in Biometric Identification Systems*, University of Pitesti, January 2013.

[3] A. Jain A., K. Nandakumar K., A. Ross: *Score Normalization in multimodal biometric systems,* Pattern Recognition, The Journal of the Pattern Recognition Society, 38 (2005).

[4] A. Ross, A. K.Jain: *Multi-Modal Biometrics: An Overview*, 12th European Signal Processing Conference (EUSIPCO), Viena, Austria, septembrie 2004

[5] V.M.Mane, D.V.Jadhav: *Review of Multimodal Biometrics: Applications, challenges and Research Areas*, International Journal of Biometrics and Bioinformatics (IJBB), Vol. 3, Issue 5, 2009.

[6] K.Bowyer, K.I.Chang: *Multi-Modal Biometrics: An Overview*, 2006.

[7] R.Snelick, M.Indovina: *Multimodal biometrics: issues in design and testing*, Proceedings of 5th International Conference on Multimodal Interfaces, Canada, 2003, pp. 68-72.

[8] A.Jain, A.Ross, S.Prabhakar: *An Introduction to Biometric Recognition*, IEEE Transaction on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, Vol. 14, Nr. 1, 2004.

[9] A. Rattani, D. R. Kisku, M. Bicego, M. Tistarelli: *Feature Level Fusion of Face and Fingerprint Biometrics*, 2007

[10] S. Soviany, S. Puşcoci, C. Soviany: *A Multi-Detector Security Architecture with Local Feature-level Fusion for Multimodal Biometrics*, Journal of Communication and Computer (JCC), no. 9/2013, David Publishing Company.