A Multimodal Biometric Identification Method for Mobile Applications Security

Sorin Soviany, Virginia Sândulescu, Sorin Puşcoci
Communications Terminals and Telematics Department
I.N.S.C.C. (Communications Studies and Research Institute)
Bucharest, Romania
sorin.soviany@inscc.ro, virginia.sandulescu@inscc.ro, sorin.puscoci@inscc.ro

Abstract – This paper presents a design approach of a reliable authentication system for mobile applications (such as those within m-Health or m-Banking areas). This means that the biometric data processing should optimize the security performance vs. the computational complexity. The security is given by the combination of fingerprint, iris and voice features that define the multimodal pattern of an individual. The complexity reduction is supported by a reduced feature space, especially for the fingerprint and iris recognition components of the overall system.

Keywords-multimodal; data fusion; biometrics; identification

I. INTRODUCTION

The multimodal biometric systems combine several human traits to improve the accuracy and security of human recognition in various real world scenarios. The multimodal design varies according to the processing stage in which the data merging is performed, before or after classification/matching, respectively [1]–[3]. The data merging requires to apply some data fusion rules either on the raw biometric data or on the matching scores or even recognition decisions. The most implemented fusion in the real applications is the post-classification one, because of its simplicity. The feature-level fusion still deals with several challenges including the features incompatibility, their homogeneity degree and the unavailability of the feature vectors in the proprietary biometric solutions. Despite of these challenges, the pre-classification fusion seems to be a very hot research and development area and a promising approach for further performance improvements.

The requirements of real world biometric security applications vary significantly with the desired security degree, complexity and costs. In the context of smartphone applications, possible scenarios in which a high security degree authentication is necessary are the applications that access sensitive, personal data. For example, in the case of m-Banking applications, end-users authenticate through their mobile devices (smartphones, tablets) in order to access and handle their bank accounts related data. Mobile medical services typically deal with personal data, like user’s medication or tracked activity. Also, smartphones become more and more powerful and so, their usage in business related activities increases, dealing with business information that also needs to be securely accessed and handled.

The mobile applications are usually resource constrained (in terms of data storage, processing and transmission speed). This justifies the strong interest in developing optimized security solutions able to efficiently meet these requirements. The biometric-based solutions for mobile applications should provide a reliable security performance degree. Some of the recent security issues that occurred with fingerprint-based smartphone authentication (like using printed replicas of registered fingerprints) justify the interest for the optimized multimodal solutions design, development and releasing on growing markets, while considering the typical requirements for various mobile applications.

In this paper, we propose a multimodal approach for biometric security solutions with focus on smartphone applications. The design embeds three biometric modalities for smartphone end-users authentication (fingerprint, iris and voice). The fusion rule is applied after the matching scores computation (post-classification). However the design allows for an update with feature-level fusion, either intra-modal (for the same biometric) and inter-modal (for several biometrics).

The paper is structured as following: Section II presents recent developments and examples of multimodal biometric systems; Section III describes the proposed multimodal system design; Section IV presents the software specification for this system; Section V concludes the paper and provides the further developments steps towards the final product.

II. RELATED WORKS

The actual developments in multimodal biometrics use various data fusion rules to combine the biometric modalities, matching scores or decisions. The multimodal biometric fusion falls in the following classes: pre-classification (sensor-level fusion, feature-level fusion) and post-classification (matching score-level fusion, decision level-fusion) [4]–[6].

There are a lot of implementations of these rules for various multi-biometric systems. For example, in [7] a multimodal system combining fingerprint and iris data is presented. The fusion design is based on a single Hamming Distance matcher that improves the accuracy of the individual biometrics.
A fusion framework for multimodal biometric authentication is presented in [8]. The model integrates face, fingerprint and iris traits into the same multimodal system using several fusion schemes: sensor-level fusion, score-level fusion and decision level-fusion. The design is based on processing several biometric modalities from different sensors using multiple processing algorithms, multiple classifiers and several fusion rules.

A bimodal biometric system with fingerprint and iris recognition is proposed in [9]. Here the fusion is applied at decision-level with the logic operator AND between decision provided by the fingerprint and iris recognition modules.

Other biometric traits have been recently integrated into multimodal solutions, for example finger vein and hand vein [10]. In this multimodal system example a score-level fusion is applied using a simple sum rule and min-max normalization.

A neural network approach for a multimodal biometric system with finger knuckle and nail is proposed in [11]. Here the fusion is applied at feature-level with a fused feature vector resulting from the concatenation of the feature sets that are extracted from each biometric. A neural network classifier is then applied on the single resulted feature vector.

The finger knuckle is integrated in another multimodal system together with the palmprint [12]. In this example, the integration is performed with a score-level fusion using a weighted sum rule to compute the overall matching score. This is used for a verification biometric recognition process. The min-max score normalization is also applied.

The integration of biometric traits of faces and both irises is proposed in [13]. Here, the three min-max normalized scores (that are computed using the Hamming Distance and Euclidian Distance) are used as input values for the Support Vector Machine (SVM) classifier with several kernels. The proposed system performance was evaluated for the following score fusion rules: the “Max” and “Min” rules, where the final score was determined by selecting the maximum and minimum input scores and using a SVM classifier with each of the biometric score as an input.

A fingerprint and iris multimodal biometric system is presented in [14]. Here, the authors apply a decision-level fusion based on fuzzy logic; each biometric component decision is weighted within the final decision making.

The multimodal biometric systems have already been implemented in real applications, sometimes in combination with other security techniques. One example is a biometric security system based on iris and palmprint recognition together with an additional personal identification number and a steganographic technique authentication [15]. In this case a feature-level fusion is applied for palmprint and iris samples. The results are used within the matching stage to generate a personal authentication number for the final ATM application.

An optimized feature-level fusion scheme is presented in [16]. The palmprint, iris and finger knuckle based features are fused using the concatenation process; then a Particle Swarm Optimization (PSO) process is applied to select the best features from the fused vector, reducing the number of useful features and therefore the recognition complexity.

III. THE MULTIMODAL IDENTIFICATION METHOD

A. The Multimodal System Architecture

The functional architecture of the multimodal system is depicted in fig. 1.

Here we have 3 Biometric data Sources (BS1, BS2, BS3) that provide the input raw data towards the further processing: BS1 for fingerprint recognition, BS2 for iris recognition, BS3 for voice recognition. The input biometric samples are images for fingerprint and iris and voice samples, respectively.

For the first two biometric modalities (fingerprint and iris) a dimensionality reduction step using a transformation algorithm like PCA (Principal Components Analysis) and LDA (Linear Discriminant Analysis) is required in order to optimize the biometric feature representation for data un-correlation (by PCA) and the best class separation (by LDA), according to these feature space transformations main properties[17].

Also, some additional feature selection algorithms could be applied for further dimensionality adjustment and for selecting the most informative features from fingerprint, data and voice. So far, these functions
define the pre-processing and feature generation stage of the overall multimodal biometric system.

The second operational component of the proposed functional architecture performs the multimodal biometric fusion of the matching scores that are separately computed for each of the integrated biometric modalities. We apply a post-classification score fusion, although the proposed architecture allows to combine the fingerprint and iris feature vectors with a feature-level (pre-classification) fusion. This remains a further research and development subject concerning the same multimodal system design with focus on mobile applications security.

B. The Feature Generation/Extraction for Fingerprint, Iris and Voice

Data processing for feature generation/extraction is performed in order to obtain the feature vectors for the further matching stage. Here, the feature generation function is designed using quite similar algorithms for fingerprint and iris components; this allows a feature-level fusion based on a functional combination of the feature vectors. This fusion will be a subject for further research.

1) Fingerprint and Iris Features

The data processing stage for fingerprint and iris features generation/extraction is based on a regional approach with textural 2\textsuperscript{nd} order statistical features using co-occurrence matrices [17]. The advantage of co-occurrence matrices is an easy adjustment of the resulted feature space dimensionality; this is typically a critical issue especially concerning the computational complexity of the biometric identification process. The overall process of feature generation for fingerprint and iris biometrics is performed in the following stages:

- manual selection of the regions of interest (ROIs) within each of the original input images (fingerprint and iris, respectively). The manual selection means that for each of the input image we apply a mask matrix to extract the most important region for the discrimination process, actually the region that contains the most useful information about the minutiae and other biometric patterns. For the moment we use a single ROI for each of the 2 biometrics, reasoned by simplicity considerations. The justification is related by the mobile applications specific in which the mobile devices have resource constraints concerning data processing, storage and transmission speed;

- the feature computing in which the 2\textsuperscript{nd} order statistical features are derived from the co-occurrence matrices. These statistical features describe the gray level distribution within the original image. Actually the procedure is applied on the previously selected ROI for both biometrics (fingerprint and iris). The resulted feature space size is easily adjusted based on the co-occurrence matrices feature extractor parameters: GLB (gray level bins number) and OFFS (displacement distance). We apply the following settings for these amounts: GLB=5 for fingerprint ROI and GLB=7 for iris ROI. This setting provides the majority of significant values for the co-occurrence matrices components, respectively less null values. This is reasoned by the requirement of informative textural features but, at the same time, a low complexity in their extraction and further processing. OFFS is an amount that accounts the pixels between the pixel pairs that are used in co-occurrence matrices computing. The main requirement for this parameter is to not exceed a given limit in order to provide a suitable pixel pairs number. Here we fix OFFS=2.

The co-occurrence matrices-based feature extraction provides the following feature space dimensionalities that represent the generated feature vectors sizes \( l(V_1) \) and \( l(V_2) \):

- for the fingerprint feature vector \( V_1 \):
  \[
  l(V_1) = (GLB_1)^2
  \]  
- for the iris feature vector \( V_2 \):
  \[
  l(V_2) = (GLB_2)^2
  \]

The resulting dimensions of the fingerprint and iris feature spaces, respectively, are the following: 25 features for fingerprint and 49 features for iris. These dimensionalities could be easily adjusted from the feature-extractor parameterization such as to generate the best co-occurrence matrices from the manually selected ROIs.

2) Voice Features

Voice processing is made in accordance to recent work of one of the authors regarding voice authentication presented in [18]. Although the current work may be particularized for both authentication and identification, the voice processing steps are the same for both usages. Relying on previous work, we know that by extracting the Mel Frequency Cepstral Coefficients (MFCC) along with the delta and delta-delta coefficients, the Linear Predicting Coding Coefficients (LPC) and voice pitch, a speaker’s voice may be described accurately enough for the intended purpose.

The MFCCs describe the signal in terms of power spectrum. The MFCCs are computed using a triangular filterbank on the spectrum computed with a Discrete Fourier Transform (DFT). Based on the MFCC coefficients, the delta (differential) coefficients and the delta-delta (acceleration) coefficients are computed. While the MFCCs describe the signal in terms of power spectrum, the delta and delta-delta coefficients describe the dynamics of the signal.

Linear predicting coding analysis is based on the fact that one frame of a voice signal can be estimated as a linear combination of the values of the previous frames.

The most common way to compute LPC is based on predicting the value of the current frame of the signal based on the previous frames and then applying the least-square method to minimize the difference between the predicted value and the real value of the signal. The coefficients are determined by minimizing the predicted value.
More detailed steps about computing the values for MFCC and LPC are described in [18].

Pitch is considered a representative feature for describing the human voice as perceived by other humans. It is also a subjective feature, meaning there is no standard formula or algorithm to compute it.

All the pitch estimators compute the fundamental frequency \( F_0 \), which is an inherent property of periodic signals and tends to correlate with the perceived pitch. \( F_0 \) represents the inverse of the smallest true period in the analyzed interval.

Voice is processed using the following steps:

- voice preprocessing: noise reduction, unvoiced segments removal;
- feature extraction: 12 MFCC plus the log energy of each frame, 13 delta coefficients, 13 delta-delta coefficients, 13 LPC coefficients and a pitch estimation, counting to a total of 53 features. The same window size may be used for extracting each of the before-mentioned features, so that the feature vector may be constructed simply by aligning the features for each moment of time.

The constructed feature vectors are used as input for a machine learning classifier. From previous work, we know that the best results in terms of accuracy, not in terms of computational power usage, were obtained by using an SVM (Support Vector Machine) classifier with a Gaussian kernel.

C. The Dimensionality Adjusting

The dimensionality adjustment by feature spaces transformations and/or feature selection is an additional and optional design and operational step that could be applied depending on the input biometric data quality and the application specific constraints. For the proposed design this step is not mandatory, however it should be considered for a further development especially if working with the feature-level fusion. Actually an initial feature space size adjustment is already done by setting the GLB and OFFS parameters of the feature extractor.

The explicit dimensionality adjustment should be applied for fingerprint and iris components with the following sequenced operations:

- PCA (Principal Component Analysis), which is an unsupervised feature space reduction technique that provides uncorrelated features [17];
- LDA (Linear Discriminant Analysis), which is a supervised feature space transform that maximizes the class separation of the input data [17];
- a feature selection procedure (such as the non-exhaustive forward or backward feature search, individual feature ranking).

This step should only be implemented if the application requirements ask for a very low dimensionality of the feature spaces, for instance less than 10 features per biometric. This would be the case if the multimodal fusion is performed at feature-level with the concatenation of the feature vectors.

For this design it only remains an optional step because the feature-level fusion will be not considered here.

D. The Matching/Classification of Biometric Samples

The matching/classification operation for fingerprint and iris is performed with a distance-based approach, also considering an identification task. This means that a matching score is computed for each enrolled identity. The system should operate for actually both biometric processes: verification and identification.

The matching score is computed using the Euclidian distance and with a normalization step in order to provide a homogeneous values between 0 and 1. Therefore for iris and fingerprint modalities the following computations are performed, given a certain feature vectors pair \((V_i, V_{i,ref})\), \(i = 1, 2\) (\(i=1\) for fingerprint feature vectors and \(i=2\) for iris feature vectors):

- The Euclidean distance \(d_i\) between the current test feature vector \((V_i)\) and the corresponding reference template \((V_{i,ref})\):

\[
d_i(V_i, V_{i,ref}) = \sqrt{\sum_{k=0}^{l(V_i)-1} (V_{i}[k] - V_{i,ref}[k])^2}, i = 1, 2
\]

where \(l(V_i)\) is the size of the feature vector \(V_i\) (the number of extracted feature for each biometric) and \(k\) is the feature index within the matching vectors;

- The score normalization with a sigmoid function that provides homogeneous scores within the range \([0, 1]\). The resulting normalized distance-based score is given by

\[
D_{i,norm}(V_i, V_{i,ref}) = \frac{1}{1 + \exp\left(-d_i(V_i, V_{i,ref})\right)}, i = 1, 2
\]

- The distance-similarity scores transformation that ensures that the highest scores show the closest similarity or matching between the current biometric pattern and the reference template. This is done by the following subtraction:

\[
S_i(V_i, V_{i,ref}) = 1 - D_{i,norm}(V_i, V_{i,ref}), i = 1, 2
\]

The voice data classification for individuals recognition is performed with a SVM classifier providing the corresponding class membership score. A multi-class extension is considered for the identification task.

E. The Multimodal Fusion Rule

For this design, we apply a matching score fusion rule. However, a further system update and development will allow to include a feature-level fusion (fingerprint and iris). This is because the design uses quite similar algorithms for feature extraction providing a certain degree of homogeneity of the resulted features.
Now we only apply the score-level fusion based on a weighted averaging of the 3 matching scores. The fusion rule is the following:

\[
S_g = \frac{\sum_{i=1}^{3} W_i \cdot S_i(V, V_{i, ref})}{\sum_{i=1}^{3} W_i}
\]

where \(S_g\) is the global score resulting from the individual scores combination, \(W_i\) is the weight that is assigned to each biometric, according to its performance.

IV. THE SOFTWARE APPLICATION DESIGN

The software application design will closely follow the processing steps presented in the previous paragraphs.

As stated in the previous chapter, the current work makes use of three biometric modalities: fingerprint, iris and voice. A software library for processing each of these types of data sources was chosen. We have searched for actively maintained open source projects that are constructed in a modular fashion, with a fairly decent documentation so that they would allow us to use and modify the code to fit our needs.

A. The Software Libraries for Biometric Data Processing

For fingerprints processing, the NIST Biometric Image Software (NBIS) [19], developed by NIST (National Institute of Science and Technology), was chosen. It is freely distributed with no licensing requirements. Written in C, NBIS is composed of replaceable modules, making it a good fit for researchers, who may adapt it to implement their own algorithms. It contains more than 50 utilities for automated fingerprint manipulation and processing, including modules for fingerprint minutiae extraction and visual representation and minutiae matching.

Iris images will be processed using a library also developed by NIST, called VASIR (Video-based Automatic System for Iris Recognition) [20]. It is currently in beta version and the source code, although incomplete, is available for download. It is written in C++, using Qt and OpenCV. It can process both video sequences and still images in different formats. It is able to reliably recognize low quality iris images. This, along with iris detection in videos, are two of the main differences between VASIR and other available iris recognition software libraries/packages, that are usually developed to work on still iris images captured in well-defined usage scenarios.

Voice processing will be accomplished using the Marsyas [21] framework for audio analysis. It is also a modular framework, allowing users and developers to use components as interchangeable building blocks. It is written entirely in C++, presenting bindings for other programming languages like Java or Python. It contains both feature extraction modules (for pitch, MFCC, LPC and others) and implementations for automatic classifiers like KNN and SVM.

B. The Integrated Multimodal Application

The integrated multimodal application will present a client-server architecture. The server side will make use of the previously described software libraries in order to implement the intense resource consuming feature extraction and matching algorithms and also the fusion rules. The end devices will only incorporate a preprocessing step for each of the biometric modality. This will reduce the stress on the end devices batteries and the need of state of the art smartphones.

V. CONCLUSIONS AND FURTHER RESEARCH/DEVELOPMENTS

The biometrics-based security techniques play an increasingly important role in securing both physical and informational systems, both in commercial and non-commercial scenarios, leading to the idea that biometric systems are going to be ubiquitous in the future.

There are multiple reliable and useful applications that may be accomplished by using parts or all the techniques and technologies described in this paper. The described work represents an ongoing research.

For the moment, we are working on developing an authentication system using the 3 presented biometric modalities. The functionalities of each of the 3 software libraries have been individually tested, but we have yet to merge them into one software project and combine the results from each of them using the proposed score fusion rule.

The next step will be the creation of a database suitable for testing the proposed system, by create virtual subjects, described by a certain voice, fingerprint and iris pattern. This will be accomplished by merging several traditional publicly available testing databases, which are designed to test each of the chosen biometric modalities on its own.

We believe that the resulting work will be a good fit both for securing real life targets and for researchers who may use our work as a starting point to develop their own multimodality biometric systems, due to the fact that the system will be composed by modules that could be easily customized. The customization allows to develop multimodal systems meeting the various real applications specific requirements.

ACKNOWLEDGEMENT

This paper is partially supported by Project PN 16170401, Contract no. 2N/09.03.2016, “Multimodal Technologies Integration for Mobile Services Security”

REFERENCES


S. Soviany, Virginia Săndulescu, Sorin Pușcoci

6


