

Face and ECG Based Multi-Modal Biometric Authentication

Ognian Boumbarov¹, Yuliyana Velcheva¹, Krasimir Tonchev¹
and Igor Paliy²

¹*Technical University of Sofia*

²*Ternopil National Economic University*

¹*Bulgaria*

²*Ukraine*

1. Introduction

A biometric system is essentially a pattern recognition system. This system measures and analyses human body physiological characteristics, such as face and facial features, fingerprints, eye, retinas, irises, voice patterns or behavioral characteristic for enrollment, verification or identification (Bolle & Pankanti, 1998). Uni-modal biometric systems have poor performance and accuracy, and over last few decades the multi-modal biometric systems have become very popular. The main objective of multi biometrics is to reduce one or more false accept rate, false reject rate and failure to enroll rate. Face Recognition (FR) is still considered as one of the most challenging problems in pattern recognition. The FR systems try to recognize the human face in video sequences as 3D object (Chang et al., 2003; 2005), in unconstrained conditions, in comparison to the early attempts of 2D frontal faces in controlled conditions. Despite the effort spent on research today there is not a single, clearly defined, solution to the problem of Face Recognition, leaving it an open question. One of the key aspects of FR is its application, which also acts as the major driving force for research in that area. The applications range from law enforcement to human-computer interactions (HCI). The systems used in these applications fall into two major categories: *systems for identification* and *systems for verification* (Abate et al., 2007). The first group attempts to identify the person in a database of faces, and extract personal information. These systems are widely used, for instance, in police departments for identifying people in criminal records. The second group finds its main application in security, for example to gain access to a building, where face is used as more convenient biometric. The more general HCI systems include not only identification or verification, but also tracking of a human in a complex environment, interpretation of human behavior and understanding of human emotions.

Another biometric modality that we use in our approach is the electrocardiogram (ECG). The modern concept for ECG personal identification is to extract the signal features using transform methods, rather than parameters in time domain (amplitudes, slopes, time intervals). The proper recognition of the extracted features and the problem of combining different biometric modalities in intelligent video surveillance systems are the novel steps that we introduce in this work.

2. Recognition of facial images

2.1 Framework for face recognition

In real case scenario, human faces often appear in scenes with complex background, rather than as a single object. In addition, they have varying appearance due to different lightning conditions, changes in pose, human expressions etc. Thus, a reliable system for FR must be robust to noise, variations and be able to work in real time. To meet these requirements we are proposing a framework for recognition of facial images depicted on Fig. 1. This framework consists of three stages, namely Face Detection (FD), Subspace Projection (SP) and Classification.

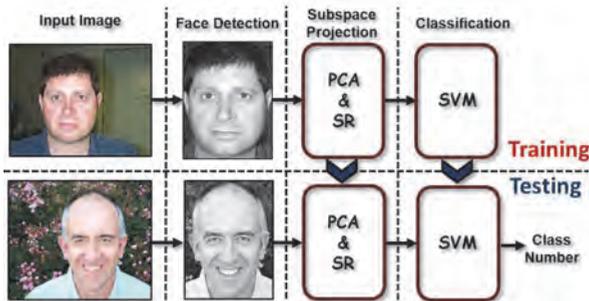


Fig. 1. Pictorial depiction of facial recognition framework

The purpose of the FD is to locate a human face in a scene and extract it as a single image. In this work, we propose a combination of two classifiers for rapid and accurate FD. The first one is faster but less precise, while the second, compensates for the imprecision of the first classifier. The second stage of the proposed framework, namely SP, is used for dimensionality reduction of the detected facial images, when represented as vectors in high-dimensional Euclidean space. Thus, it is necessary to transform them from the original high-dimensional space to a low dimensional one for alleviating the *curse of dimensionality*. The SP is based on Principal Component Analysis (PCA) and Spectral Regression (SR) algorithms. The PCA discovers the subspace which contains all vectors of facial images and we use it mainly to remove noises. PCA also preserves Euclidean distances between vector pairs in the subspace. Based on this, further dimensionality reduction is done by using the SR algorithm. This algorithm is robust with respect to the variation of lightning conditions and human expressions. Finally we perform classification using Support Vector Machines classifier in the subspace. In the following, the three stages of the proposed FR framework we will be discussed in details.

2.2 Face detection

A combination of rapid cascaded classifier and accurate monolithic one is used as a two level Face Detection algorithm. The first level is represented by the Haar-like features' cascade of weak classifiers, which is responsible for fast detection of face-like objects. The second level is a Convolutional Neural Network (CNN) used for filtration of falsely detected faces. The Haar-like features' cascade of weak classifiers allows detecting face candidates very quickly. It consists of a cascade of one or more weak classifiers. The weak classifier's input is represented by Haar-like feature with a value (Viola & Jones, 2004):

$$Feat(x) = s_w \times SUM_w + s_b \times SUM_b \quad (1)$$

where x is input image's sub-window, s_w and s_b - whole rectangle's and its black part's weights accordingly, SUM_w and SUM_b - whole rectangle's and its black part's sums of pixels. A weak classifier's output value is:

$$h(x) = \begin{cases} 1, & \text{if } Feat(x) \leq \Theta, \\ -1, & \text{if } Feat(x) > \Theta, \end{cases} \quad (2)$$

where Θ - weak classifier's threshold. The cascade of weak classifiers is a linear combination of weak classifiers (Viola & Jones, 2004):

$$H(x) = \sum_t^T \eta_t \times h_t(x), \quad (3)$$

where T - weak classifiers' number, η_t - t -weak classifier's weight. The AdaBoost algorithm (Freund & Schapire, 1997) is used for training of the cascade of weak classifiers and the selection of the most important Haar-like features. The second level uses the Convolutional Neural Network (Lecun et al., 1998) which is more robust to variations of the input image, compared to other known classifiers. The output value of a neuron with bipolar sigmoid transfer function and with the coordinates (m, n) of p -plane and l -layer is (Kurylyak et al., 2009):

$$y_{m,n}^{i,p} = \frac{2}{1 + \exp(-WSUM_{m,n}^{i,p}(x))}, \quad (4)$$

where x is input face candidate's image, $WSUM$ - neuron's weighted sum calculated by (Kurylyak et al., 2009):

$$WSUM_{m,n}^{i,p}(x) = \left(\sum_{k=0}^{K-1} \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} y_{2m+r,2n+c}^{l-1,k} \times w_{r,c}^{l,p,k} \right) - b^{l,p}. \quad (5)$$

Here K is input planes' number (as well as convolutional kernels), R and C are convolutional kernel's height and width, $w_{r,c}^{l,p,k}$ is synaptic weight with coordinates (r, c) in the convolutional kernel between k -plane of the $(l-1)$ -layer and p -plane of the l -layer, $b^{l,p}$ is neurons' bias of the p -plane and l -layer. The CNN uses a sparse structure instead of a fully-connected one; also its number of layers is decreased. In order to increase the neural network's processing speed, convolution and subsampling operations are performed in each plane simultaneously (Simard et al., 2003) Fig. 2.

2.3 Facial features extraction with subspace projection

2.3.1 Principal Component Analysis

Principal Component Analysis (PCA) is a very popular method for dimensionality reduction in machine learning and statistics communities. It can be considered as method learning the basis vectors spanning the linear subspace, called *principal subspace*, containing all data points embedded in that subspace. This basis is determined by the non-zero eigenvectors of data covariance matrix, and the dimension of the subspace is their number. Usually this number is much less than the dimension of the ambient space and dimensionality reduction is performed by projection the data points onto that subspace.

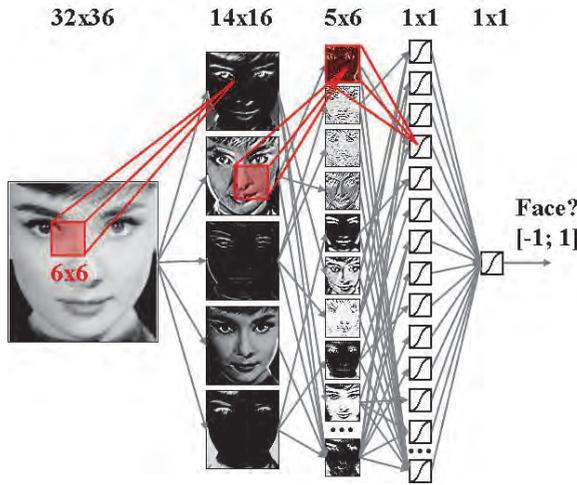


Fig. 2. Convolutional neural network's structure for the combined cascade of neural network classifiers

PCA can be formulated in two ways (Bishop, 2007). The first one is called *minimum-error formulation* and the target is to minimize the mean squared error between the data and its projections onto the subspace. In the second approach, called *maximum variance formulation*, the goal is to maximize the variance of the projected data. In this work we approach PCA by the former one. Let $\{\mathbf{x}_i\}_{i=1}^M \subset \mathbb{R}^N$ be a data set of measurements of physical phenomenon, drawn from an unknown probability distribution. We seek a subspace of dimension $D \ll N$ such that the variance of the projected data onto it is maximized. First step of PCA involves computation of data covariance matrix by:

$$\mathbf{C} = \sum_{i=1}^M (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T, \quad (6)$$

where $\bar{\mathbf{x}} = \sum_{i=1}^M \mathbf{x}_i$ is the mean vector. Then the variance of the projected data onto single direction $\mathbf{v} \in \mathbb{R}^N$ is:

$$\frac{1}{N} \sum_{i=1}^N (\mathbf{v}^T \mathbf{x}_i - \mathbf{v}^T \bar{\mathbf{x}})^2 = \mathbf{v}^T \mathbf{C} \mathbf{v}. \quad (7)$$

Maximizing the variance, means maximizing the quadratic term $\mathbf{v}^T \mathbf{C} \mathbf{v}$. But when $\|\mathbf{v}\| \rightarrow \infty$ it follows that $|\mathbf{v}^T \mathbf{C} \mathbf{v}| \rightarrow \infty$, hence it is desired to constrain the problem in appropriate way. This is achieved by the normalized condition $\mathbf{v}^T \mathbf{v} = 1$. By introducing Lagrange multipliers we reformulate the problem from constrained to unconstrained maximization one:

$$\mathbf{v}^T \mathbf{C} \mathbf{v} + \lambda (1 - \mathbf{v}^T \mathbf{v}) \rightarrow \max, \quad (8)$$

where λ is a Lagrange multiplier. Taking the partial derivatives of 8 with respect to \mathbf{v} and setting to zero, we end up with the eigenvalue problem $\mathbf{C} \mathbf{v} = \lambda \mathbf{v}$. Thus the direction defining maximum variance of projected data is the leading eigenvector called *principal component*. Additional principal component can be found by calculating new direction which maximizes

8 and it is orthogonal to the previous ones considered. By arranging the eigenvalues on the diagonal of the diagonal matrix Λ and the eigenvectors as e columns of the matrix \mathbf{V} we can write the eigenvalue problem as $\mathbf{CV} = \Lambda\mathbf{V}$. By selecting only D eigenvectors of \mathbf{V} corresponding to the leading eigenvalues and arranging them in matrix $\tilde{\mathbf{V}}$ the subspace projection is defined by $\tilde{\mathbf{x}}_i = \mathbf{x}_i^T \tilde{\mathbf{V}}$ for $i=1 \dots M$.

2.3.2 Spectral regression framework for dimensionality reduction

Dimensionality reduction can be well interpreted in a *graph embedding* fashion. Such interpretation is very intuitive and also opens the possibility of developing a new approaches to the problem of pattern recognition. Advantage of graph embedding is the unifying power to the most of the dimensionality reduction methods such as Linear Discriminant Analysis (LDA), Locally Linear Embeddings (LLE), Locality Preserving Projections (LPP), ISOMAP as proposed in (Yan et al., 2007). Consider a data set $\{\mathbf{x}_i\}_{i=1}^M \subset \mathbb{R}^N$ represented as points in Euclidean space. Each data point can be viewed as e vertex of adjacency graph $\Gamma = \{X, W\}$ with edges W defined under some rule. Depending on the rule, graph edges can be distance, or a measure of similarity, between data points (vertices of Γ). The *graph Laplacian* is defined by $L = D - W$, where $D_{ii} = \sum_{j \neq i} W_{ij}$. The goal of graph embedding is to find a low dimensional representation of each vertex, while preserving similarities between the vertex pairs. This is achieved by minimizing (Chung, 1997):

$$\sum_{ij} (y_i - y_j)^2 W_{ij} = 2\mathbf{y}^T L \mathbf{y}, \quad (9)$$

where $\mathbf{y} = [y_1, y_2, \dots, y_M]^T$ is a map of graph vertices (data points) on the real line. If we select a linear map of the form $y_i = \mathbf{a}^T \mathbf{x}_i$, the problem 9 can be reduced to:

$$\mathbf{a}^* = \arg \min_{\mathbf{a}} \frac{\mathbf{a}^T X W X^T \mathbf{a}}{\mathbf{a}^T X D X^T \mathbf{a}}, \quad (10)$$

where $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M]$ is the data matrix. This approach is called *Linear extension of graph embedding*, and the optimal solution of 10 can be found by the *generalized eigenvalue problem*:

$$X W X^T \mathbf{a} = \lambda X D X^T \mathbf{a}. \quad (11)$$

With different choices of W , 11 can be formulated as LDA, LPP etc. Solving 11 can be computationally intensive in cases where W is a dense matrix of very high dimensions. To overcome this issue in (Cai et al., 2007) proposed a method, called Spectral Regression, which casts solving the generalized eigenvalue problem 11 into regression framework. The advantages of this approach are: (1) Solving a regression problem less computationally intensive than eigenvalue problem; (2) Regression can be solved with regularization term controlling the complexity and avoid overfitting; (3) By selecting regression terms, various properties can be achieved, such as sparsity. Furthermore, if we chose a function in a Reproducing Kernel Hilbert Space $y_i = \sum_{j=1}^M \alpha_j K(\mathbf{x}_j, \mathbf{x}_i)$, where $K(\cdot, \cdot)$ is a Mercer kernel, SR can be extended in kernel mode.

Algorithmically SR is performed as follows (Cai, 2009):

1. Construct the weights matrix W :

- Set $W_{ij} = 1/l_k$ if \mathbf{x}_i and \mathbf{x}_j both belong to the k -th class where l_k is the number of samples in it;

- Set $W_{ij} = \delta \cdot s(i, j)$ if \mathbf{x}_i is among the p -nearest neighbor of \mathbf{x}_j or vice-versa. The parameter δ is used to adjust the weight between supervised and unsupervised neighbor information. $s(i, j)$ is a function evaluating the similarity between \mathbf{x}_i and \mathbf{x}_j .

This function can be the heat kernel function $s(i, j) = \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{2\sigma})$ or simple minded $s(i, j) = 1$, where $\sigma \in \mathbb{R}$;

- Set $W_{ij} = 0$ otherwise.
2. **Responses Generation:** Solve the eigenvalue problem $\mathbf{W}\mathbf{y} = \lambda\mathbf{D}\mathbf{y}$ and select the K largest eigenvectors $\{\mathbf{y}_k\}_{k=1}^K$, where \mathbf{D} is diagonal matrix with elements on the diagonal $D_{jj} = \sum_i W_{ij}$.
 3. **Regression:** Ridge regression solves the quadratic regression problem with Euclidean norm penalty:

$$\mathbf{a}_k = \arg \min_{\mathbf{a}} \left(\sum_{i=1}^M (\mathbf{a}^T \mathbf{x}_i - y_i^k)^2 + \alpha \|\mathbf{a}\|^2 \right) \quad (12)$$

In Lasso regression, the regression problem is solved with L_1 -norm penalty:

$$\mathbf{a}_k = \arg \min_{\mathbf{a}} \left(\sum_{i=1}^M (\mathbf{a}^T \mathbf{x}_i - y_i^k)^2 + \alpha \sum_{j=1}^N |a_j| \right) \quad (13)$$

In both 12 and 13 the solutions are represented by $\{\mathbf{a}_k\}_{k=1}^K \subset \mathbb{R}^N$.

4. **Subspace Projection:** Perform dimensionality reduction by projecting on a lower dimensional space $\mathbf{x} \rightarrow \mathbf{z} = \mathbf{A}^T \mathbf{x}$, where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_K]$ is matrix of the solution vectors from the regression step.

For the purpose of this work, we select two different regularization terms in the regression step. The first one, called Ridge, is the standard quadratic regression 12. The dimensionality of the output space in this case is $c - 1$ and $K = c$. The second mode, called Lasso, uses L_1 -norm for the regularization term which induces sparsity on $\{\mathbf{a}_k\}_{k=1}^K$. We test the proposed framework for face recognition with both modes of SR.

2.4 Classification of facial images

Support Vector Machine is a supervised learning algorithm used for classification in two classes. The aim of SVM is to find a N -dimensional hyperplane that optimally separates the data. Optimally in this case means that the margin, between nearest data points and the hyperplane, will be maximized. Unfortunately in real problems data is rarely separable by a hyperplane but can be separated by a non-linear surface. SVM can be transformed to a non-linear classifier by applying the kernel trick (Vapnik, 1998). This way data is mapped implicitly in a higher dimensional space where it can be separated by a hyperplane. Let $\{\mathbf{x}_i\}_{i=1}^M \subset \mathbb{R}^N$ be a data set with class labels vector $\mathbf{y} = [y_1, y_2, \dots, y_M]$ such that $y_i \in \{-1, 1\}$, $i = 1, 2, \dots, M$. Learning the parameters of the support vectors is performed by solving the constrained optimization problem:

$$\begin{aligned}
\boldsymbol{\alpha}^* &= \arg \min_{\boldsymbol{\alpha}} \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{Q} \boldsymbol{\alpha} - \mathbf{e}^T \boldsymbol{\alpha} \\
\text{s.t. } &\mathbf{y}^T \boldsymbol{\alpha} = 0, \\
&0 \leq \alpha_i \leq C, i = 1, 2, \dots, M,
\end{aligned} \tag{14}$$

where $C > 0$ is an upper bound, \mathbf{e} is a vector with unit elements, \mathbf{Q} is $M \times M$ positive semi-definite matrix defined by $Q_{ij} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$ and $K(\cdot, \cdot)$ is a Mercer kernel (Vapnik, 1998). The decision function for unknown sample \mathbf{x} is given by:

$$y^* = \text{sgn} \left(\sum_{i=1}^M \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b \right). \tag{15}$$

2.5 Experimental testing and results

2.5.1 Data set

We apply the proposed framework on a part of the face image database from the Computational Vision at the California Institute of Technology, USA (Caltech-CV-Group, 1999). The original database contains JPEG images of faces of 19 persons with different lighting/expressions/backgrounds/ and male or female. We select prepare a subset of the database with 10 images per person. Next we split this subset in two groups with even images per class, i.e. 5 images per person for the first and second groups.

2.5.2 Experimental setup and testing

For each group of the subset we perform training followed by testing with the remaining group and for each run we calculate the recognition rate. The final recognition rate is calculated by averaging the rates of each run. With this protocol we test our framework with 'Ridge' and 'Lasso' regression setting of SR, where the former is tested with different value of sparsity. The dimensionality of the subspace for 'Ridge' and 'Lasso' is 18 and 30 respectively. For the first one it is controlled by the number of classes and for the second one it is determined by experiments. In Table 1 are displayed the results of our experiments.

Method	Accuracy, %
SR Ridge	94.21
SR Lasso (30)	38.94
SR Lasso (50)	73.15
SR Lasso (80)	95.26

Table 1. Recognition rate for FR framework setups

Controlling the sparsity parameter over 80 is not possible because it is bounded above by the cardinality of the training set which is 95 in our case. With value 80 of sparsity parameter the proposed framework achieves best recognition rate as results reveal. For this reason, we select Lasso mode of SR for combining with ECG biometric modality.

3. ECG personal identification

3.1 State of art

The ECG is widely used as diagnostic tool in the cardiology because of its clinical significance and considering its noninvasive nature. It is a registration of the electrical activity of the

human heart over time. The genesis of this electrical activity is discussed in detail in (Malmivuo & Plonsey, 1995). The registration of ECG is performed by measuring the generated electrical voltage between pair of leads attached on the human body according to defined standards. Usually more than 2 electrodes are used and the standard for clinical electrocardiography is an ECG acquisition taken from 12 channels, thus "viewing" the electrical activity from 12 different angles. According to electrodes placement, the leads can be divided into two groups: limb leads (RA, LA, LL) and augmented leads (V1-6) (Fig. 3). Each of these groups represents the electric field of the heart in a given plane (frontal or horizontal respectively).

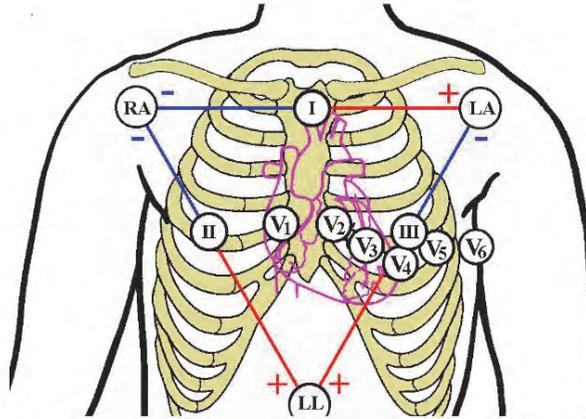


Fig. 3. Electrodes placement in standard 12 channel electrocardiography

Obviously the anatomical differences in heart muscle among individuals can be "seen" better in ECG when using all 12 channels. This type of acquisition is very impractical for real world application, so it is important to extract appropriate features when using only easy accessible leads, for example left and right arm.

A typical "healthy" ECG waveform has three distinct regions called waves (Malmivuo & Plonsey, 1995) (Fig. 4), however this morphology strongly depends on used leads, patient's condition, etc. A cardiac cycle starts with P wave, which represents the depolarization (electrical discharge) of the heart's atria. The QRS complex is a transient signal deflection related to the depolarization of the ventricles. The cardiac cycle ends with repolarization (electrical recharge) of the ventricles, which is seen in ECG as T wave. The remaining regions are referred as baseline.

The researches on the use of ECG as modality for personal identification have started since the beginning of the 21-st century. The first published articles covering this field are concentrated mainly in proving the personal discriminative characteristics of ECG as well as theirs relatively time invariance. In (Biel et al., 2001) are studied the time domain characteristics of such signals taken from 20 subjects (Fig. 4). In the cited article time domain features were reduced down to 7 according to achieved results from Principal Component Analysis (PCA). It was proven by experiments that it is possible to identify a person using ECG taken from only one lead.

In (Israel et al., 2005) can be seen a similar approach for ECG features extraction. The time domain characteristics are referred in this article as analytic features. The signal is

preprocessed in terms of bandpass filtration. After this, the peaks of P wave, R wave and T wave are found as local maxima determination is a given region. The minimum radius curvature is incorporated to find the onsets and offsets of such waves. The relevant features are selected using Wilks' Lambda method.

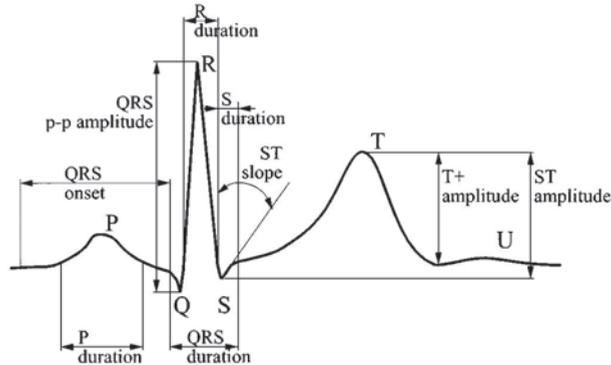


Fig. 4. Graphical representation of time domain ECG characteristics used as features for ECG personal identification (Biel et al., 2001)

Use of analytic ECG features has some disadvantages. Firstly, it requires sophisticated methods for automated ECG segmentation. Secondly, the amplitude parameters of the ECG depend not only on the unique anatomic characteristics of the human heart but also on electrodes contact, electrodes placement, etc. Finally, when the signal is highly influenced by noise, automated determination of analytic features with acceptable accuracy could fail.

As described above, the ECG segmentation is a primary and unavoidable process for any kind automated technique for features extraction. In order to avoid the need of precise determination of ECG wave boundaries, in (Plataniotis et al., 2006) is described an original approach. Firstly, the signal is divided into subsets in way that the subsets contain at least two complete cardiac cycles. Secondly, the normalized Autocorrelation Function (ACF) is calculated for a subset. Finally, the Discrete Cosine Transform (DCT) is applied on the ACF. About 40 of the most significant DCT coefficients serve as features for the classification process. The motivation behind using the ACF is the capability of non random patterns detection in the signal.

3.2 Methods for automated ECG segmentation

ECG segmentation is a key process in automated ECG identification systems. In healthy persons the ECG has strong cyclic recurrence. This is exploited in numerous approaches for full automated segmentation using Hidden Markov Models (HMM) (Hughes et al., 2004), (Boumbarov et al., 2009).

When dealing with morphological features, the full segmentation is not necessary to be performed. Instead of that some simple, fast and still effective ways can be used to extract the subsets consisting all waves and complexes in a complete cardiac cycle. The approach in (Velchev, 2010) proposes an identification of R peaks based on morphological filtration and histogram analysis. The morphological filtration is an interaction between the signal $x(t)$ and a predefined simple shape $g(t)$ called structuring element. The domains $x(t)$ and $g(t)$ are respectively X and G . The morphological filtration is based on two simple operations: dilatation $D(x, g)(t)$ and erosion $E(x, g)(t)$. They are defined as:

$$D(x, g)(y) = \max_{t-u \in X, u \in G} (x(t-u) + g(u)) \quad (16)$$

and

$$E(x, g)(y) = \min_{t-u \in X, u \in G} (x(t-u) - g(u)). \quad (17)$$

The morphological operations *Close* and *Open* are defined as follows:

$$Close(x, g) = E(D(x, g), g) \quad (18)$$

and

$$Open(x, g) = D(E(x, g), g). \quad (19)$$

The operation $Open(x, g)$ smoothers the convex peaks of the signal, while $Close(x, g)$ smoothers the concave peaks.

The structure of morphological filter for QRS complexes exaggeration is according to the following relation:

$$x'(t) = x(t) - \frac{1}{2}(Open(Close(x, g), g) + Close(Open(x, g), g)), \quad (20)$$

where $x'(t)$ is the filtered signal. The chosen structuring element is "disk". Determining its radius is a highly empirical procedure. However, it is clear that the diameter of the "disk" should be less than wf_s , where w is the smallest width of the P and T waves and f_s is the sampling frequency. In Fig. 5 are shown two examples achieved according to 20 and using the "disk" as structuring element. As can be seen it is possible to choose the parameter of the structuring element in such way, the QRS complexes are remaining as much close to the original while the P and T waves are considerably suppressed.

The next procedure is to find an appropriate threshold for coarse detection of the R peaks. A histogram is calculated for each detected local maxima of the signal. This histogram for most of the cases will be bimodal. The optimal threshold is calculated according to Otsu's method (Otsu, 1979).

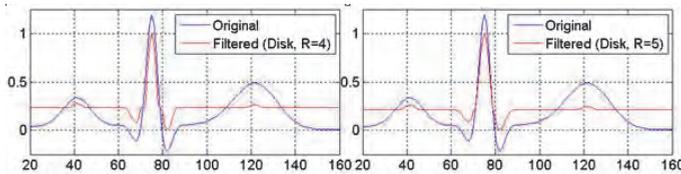


Fig. 5. Results from morphological filtration of ECG signal with structuring element 'disk' with radius 4 (left) and 5 (right). The sampling frequency is 256Hz

The last procedure in the segmentation process is to find a sample in the baseline between T and next P wave. This sample has to belong to the smoothest part of the region and its determination doesn't have to be extremely precise. Let $\mathbf{z} = [z_1, z_2, \dots, z_{N-1}]$ be a vector which elements are the numbers of the wanted samples and N is the number of the identified R peaks. An element of \mathbf{z} is calculated according to:

$$z_i = \arg \min_{\tau} \left(\sqrt{\left(CW_x^{\psi_{\text{gauss1}}}(s, \tau) \right)^2 + \left(CW_x^{\psi_{\text{gauss2}}}(s, \tau) \right)^2} \right), \quad (21)$$

$$\tau \in [r_i + (\Theta_2 - \Theta_1)(r_{i+1} - r_i)],$$

where $CW_x^{\psi_{\text{gauss1}}}$ and $CW_x^{\psi_{\text{gauss2}}}$ are the continuous wavelet transforms of x with wavelet function first and second derivative of the Gaussian, s is the scale of the wavelet transform, τ is the translation, r_i is the position of the i -th detected R peak and $\Theta_2 > \Theta_1$ are time values measured from R peak position between which the wanted sample is expected to be found. The values of s should be chosen large enough in terms for noise robustness and small enough in terms to achieve an accepted accuracy of identification. In Fig. 6 is shown an example result using the described approach.

As a final result from these procedures the ECG signal is segmented into complete cardiac cycles containing all important information in the middle of the resulting segments. Additionally the proposed approach doesn't require any supervised training.

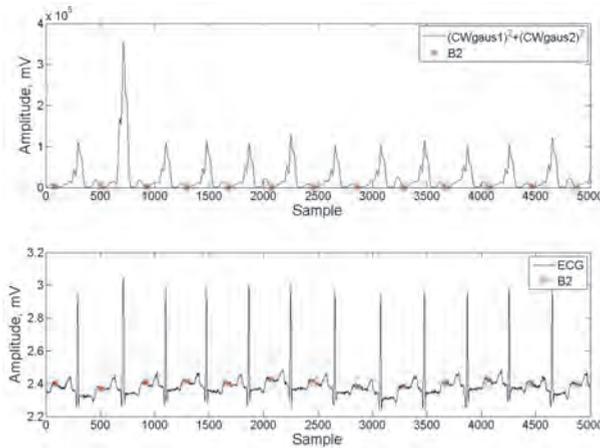


Fig. 6. An example of detection of samples belonging to the baseline between T and next P wave using combination of wavelet transforms with first and second derivative of the Gaussian as wavelet functions

3.3 ECG features extraction using linear and nonlinear projections in subspaces

As mentioned above, using the time domain characteristics as unique ECG features for personal identification has many significant drawbacks. An alternative approach is extraction of morphological features from ECG. These features could be extracted from a whole cardiac cycle in ECG, thus the need of full segmentation is eliminated. In this sense these features can be considered as holistic. They consist simultaneously amplitude and temporal characteristics of the ECG waves as well as their shape.

In this section two approaches for holistic features extraction are described. The first is based on linear projections in subspaces: PCA and Linear Discriminant Analysis (LDA). The second uses nonlinear versions of PCA and LDA: Kernel Principal Component Analysis

(KPCA) and Generalized Discriminant Analysis (GDA). A block diagram for an ECG personal identification system based on described features is shown in Fig. 7.

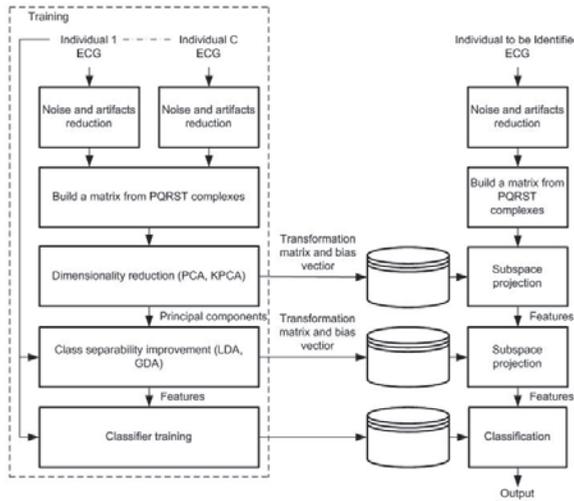


Fig. 7. Block diagram for ECG personal identification system based on features extracted using linear or nonlinear projections in subspaces

3.3.1 ECG features extraction using linear projections in subspaces

PCA is a statistical technique for dimensionality reduction of high dimensional correlated data down to a very few coefficients called principal components (Castells et al., 2007). This technique is optimal in terms of retaining as small as possible the Mean Squared Error (MSE) between the original signal and the restored signal from reduced set of principal components. Let the ECG signal is automatically segmented into PQRST complexes. These complexes are aligned and arranged as rows in a matrix. In general they differ in their length. The number of columns of the matrix is chosen larger than the maximal expected length of the PQRST complexes. The PQRST complexes are arranged in the matrix in such way the R peaks are in the middle of rows (Fig. 8). The elements in each row before and after the copied PQRST complex are set as same as the first and last sample of the complex respectively.

The training procedure requires a common training matrix $\tilde{\mathbf{X}} = [\tilde{x}_{ij}]_{N \times M}$ built from PQRST complexes taken from all individuals. The normalized training matrix $\tilde{\mathbf{Z}} = [\tilde{z}_{ij}]_{N \times M}$ is calculated according to:

$$\tilde{\mathbf{Z}} = [\tilde{z}_{ij}]_{N \times M} = \frac{1}{\mathbf{u}\tilde{\boldsymbol{\sigma}}} (\tilde{\mathbf{X}} - \mathbf{u}\tilde{\boldsymbol{\mu}}), \quad (22)$$

where \mathbf{u} is a column vector of ones with number of elements N , $\tilde{\boldsymbol{\mu}} = [\tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_M]$ is the mean vector (bias vector) and $\tilde{\boldsymbol{\sigma}} = [\tilde{\sigma}_1, \tilde{\sigma}_2, \dots, \tilde{\sigma}_M]$ is the vector of the standard deviation of $\tilde{\mathbf{X}}$. The covariance matrix $\boldsymbol{\Sigma}_{\tilde{\mathbf{Z}}} = [\tilde{\sigma}_{\tilde{z},ij}]_{M \times M}$ of $\tilde{\mathbf{Z}}$ is calculated according to:

$$\boldsymbol{\Sigma}_{\tilde{\mathbf{Z}}} = E \left\{ \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T \right\} = \frac{1}{N-1} \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T, \quad (23)$$

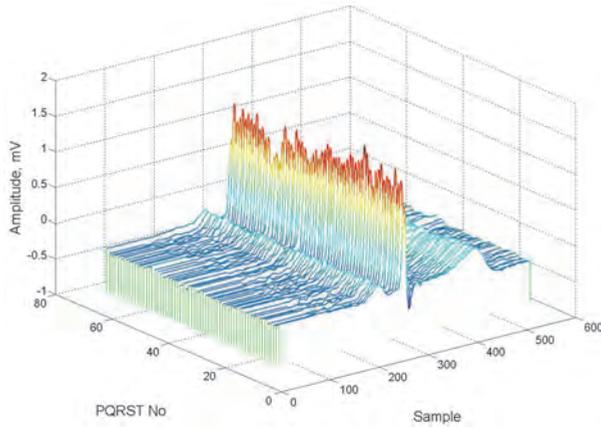


Fig. 8. Aligned PQRST complexes form ECG signal

where $E \{ \}$ denotes expectation.

The next procedure is called eigendecomposition of $\Sigma_{\hat{\mathbf{Z}}}$ according to the relation:

$$\Sigma_{\hat{\mathbf{Z}}} \mathbf{V} = \mathbf{V} \mathbf{\Lambda}, \quad (24)$$

where $\mathbf{\Lambda} = [\tilde{\lambda}_{ij}]_{M \times M}$ is a diagonal matrix and the values in its main diagonal are the eigenvalues of $\Sigma_{\hat{\mathbf{Z}}}$. The columns of $\mathbf{V} = [\tilde{v}_{ij}]_{M \times M}$ are the eigenvectors of $\Sigma_{\hat{\mathbf{Z}}}$.

The eigenvalues in $\mathbf{\Lambda}$ and the columns of \mathbf{V} have to be rearranged in descending order of the eigenvalues. The number of principal components is equal to the original length of the input observations M so they have to be reduced. The reduced number of principal components L is calculated according to:

$$L = \min_i \left(mse(\hat{\mathbf{Z}})_i \leq mse_{tr} \right), \quad (25)$$

where $\mathbf{mse}(\hat{\mathbf{Z}}) = [mse(\hat{\mathbf{Z}})_1, mse(\hat{\mathbf{Z}})_2, \dots, mse(\hat{\mathbf{Z}})_M]$ is the vector from MSE values between original matrix $\hat{\mathbf{Z}}$ and the matrix $\hat{\mathbf{Z}}$, which is restored back from reduced set of principal components and mse_{tr} is a preliminary determined value. The elements of $\mathbf{mse}(\hat{\mathbf{Z}})$ are calculated according to (Sanei & Chambers, 2007):

$$\mathbf{mse}(\hat{\mathbf{Z}})_i = \sum_{j=1}^M \lambda_{jj} - \sum_{k=1}^i \lambda_{kk}, \quad i = 1, \dots, M. \quad (26)$$

The transformation matrix $\mathbf{W} = [w_{ij}]_{M \times L}$ is built as:

$$w_{ij} = v_{ij}, \quad i = 1, \dots, M. \quad (27)$$

In the process of authentication the features \mathbf{Y} of the authenticated individual are calculated as (Franc & Hlavac, 2004):

$$\mathbf{y} = \left(\mathbf{W}^T \left(\frac{1}{\mathbf{u}\tilde{\sigma}} (\mathbf{X} - \mathbf{u}\tilde{\mu}) \right) \right), \quad (28)$$

where \mathbf{u} is a column vector of ones with number of elements N . Typically in the set of PQRST complexes there exist an amount of non specific information. This is due to the presence of noise and artifacts or can be caused by an arrhythmia. These non specific PQRST complexes should be excluded from the analysis. A convenient way is to use the Hotelling statistics $\hat{\mathbf{T}}^2 = [\hat{t}_1^2, \hat{t}_2^2, \dots, \hat{t}_N^2]$. A given element indicates how far is the training sample from the centroid of the formed cluster (Hotelling, 1931):

$$\hat{t}_i^2 = \sum_{j=1}^L \frac{\hat{y}_{ij}^2}{\lambda_j}, i = 1, \dots, N. \quad (29)$$

The criterion for excluding a given PQRST complex from the analysis is $\hat{t}_i^2 > \hat{t}_{tr}^2$, where the threshold value \hat{t}_{tr}^2 is:

$$\hat{t}_{tr}^2 = \frac{1}{N} \sum_{i=1}^N \hat{t}_i^2 + \sqrt{\frac{1}{N-1} \sum_{j=1}^N \left(\hat{t}_j^2 - \frac{1}{N} \sum_{i=1}^N \hat{t}_i^2 \right)} \quad (30)$$

3.3.2 ECG features extraction using linear projections in subspaces

In Fig. 9 is given an example of features distribution of five individuals using two different techniques - PCA and KPCA. As can be seen the extracted features using PCA aren't linearly separable. Despite the complications of the process, the results from KPCA are much better.

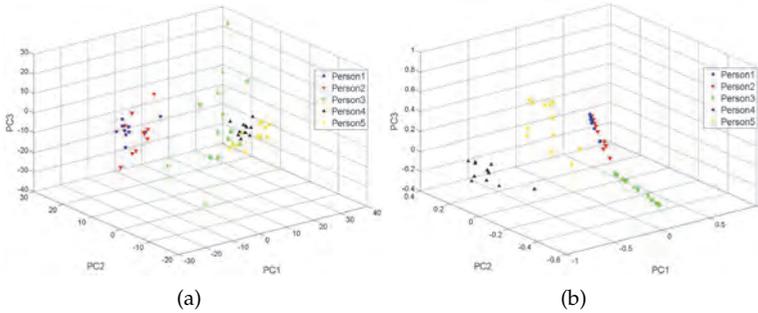


Fig. 9. Distribution of the first three principal components from ECG signals taken from 5 individuals using PCA - (a) and KPCA - (b)

According to (Schölkopf et al., 1998) KPCA could be considered as performing PCA in other (usually higher dimensional) space \mathcal{F} . The mapping of the original input vector (PQRST complex) in the new space is expressed as $\mathbf{x} \rightarrow \Phi(\mathbf{x})$. Let $\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_M)$ be the projections of the input vectors, where M is the number of observations. An input vector is expressed as $\mathbf{x}_j = [x_{j,1}, x_{j,2}, \dots, x_{j,N}]$, where N is the original dimensionality. The key process behind KPCA is to perform an eigendecomposition of the covariance matrix $\Sigma_{\Phi(\mathbf{x})}$:

$$\lambda \mathbf{v} = \Sigma_{\Phi(\mathbf{x})} \mathbf{v}, \quad (31)$$

where \mathbf{v} is an eigenvector of $\Sigma_{\Phi(\mathbf{x})}\mathbf{v}$, λ is the corresponding eigenvalue and $\Sigma_{\Phi(\mathbf{x})} = \frac{1}{M} \sum_{j=1}^M \Phi(\mathbf{x}_j)\Phi(\mathbf{x}_j)^T$. Taking into account the definition of $\Sigma_{\Phi(\mathbf{x})}$, \mathbf{v} is a linear combination of the vectors $\Phi(\mathbf{x}_j)$, $j = 1, \dots, M$:

$$\mathbf{v} = \sum_{j=1}^M w_j \Phi(\mathbf{x}_j). \quad (32)$$

Using 31 and 32 we obtain:

$$\lambda \sum_{j=1}^M w_j \Phi(\mathbf{x}_j) = \frac{1}{M} \sum_{j,k=1}^M w_j \Phi(\mathbf{x}_k) \left(\Phi(\mathbf{x}_k)^T \Phi(\mathbf{x}_j) \right) \quad (33)$$

which is equivalent of system of M equations:

$$\begin{aligned} \lambda \sum_{j=1}^M w_j \left(\Phi(\mathbf{x}_k)^T \Phi(\mathbf{x}_j) \right) &= \\ &= \frac{1}{M} \sum_{j,k=1}^M w_j \left(\Phi(\mathbf{x}_k)^T \Phi(\mathbf{x}_j) \right) \left(\Phi(\mathbf{x}_k)^T \Phi(\mathbf{x}_j) \right), \forall l = 1, \dots, M. \end{aligned} \quad (34)$$

In the last expression the inner products in the new space give the possibility not to deal directly with Φ , but to use the kernel matrix $\mathbf{K}[k_{ij}]_{M \times M}$:

$$k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle, i, j = 1 \dots M. \quad (35)$$

where the operator $\langle \cdot, \cdot \rangle$ stands for inner product and k is the kernel function. Using kernel matrix 34 is transformed into:

$$M\lambda \mathbf{K}\mathbf{w} = \mathbf{K}^2\mathbf{w}, \quad (36)$$

where $\mathbf{w} = [w_1, w_2, \dots, w_M]^T$. The last is equal to (Schölkopf et al., 1998):

$$M\lambda \mathbf{K}\mathbf{w} = \mathbf{K}\mathbf{w}, \quad (37)$$

Determining \mathbf{w} for each principal component is calculating the eigenvectors of \mathbf{K} . Let $\lambda_1 \leq \lambda_2 \leq \dots, \lambda_M$ is the full set of arranged eigenvalues, $\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^M$ is the set of eigenvectors and λ_p is the first nonzero eigenvalue. According to (Schölkopf et al., 1998) $\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^M$ have to be normalized in the way $\langle \mathbf{v}^k, \mathbf{v}^k \rangle = 1, \forall k = p, \dots, M$. Using 37 the normalization is expressed as:

$$\begin{aligned} \sum_{i,j=1}^M w_i^k, w_j^k \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle &= \sum_{i,j=1}^M w_i^k, w_j^k k_{ij} \\ &= \langle \mathbf{w}^k, \mathbf{K}\mathbf{w}^k \rangle = \lambda_k \langle \mathbf{w}^k, \mathbf{w}^k \rangle = 1, \forall k = p, \dots, M \end{aligned} \quad (38)$$

Obtaining the projection of x onto its principal components subspace would require calculating the projections of the eigenvectors $\mathbf{v}^k, k = p, \dots, M$ onto \mathcal{F} :

$$\langle \mathbf{v}^k, \Phi(\mathbf{x}) \rangle = \sum_{j=1}^M w_j^k \langle \Phi(\mathbf{x}_j), \Phi(\mathbf{x}) \rangle \quad (39)$$

A traditional approach for improving the class separability is the Linear Discriminant Analysis (LDA) (Theodoridis & Koutroumbas, 2006), however if the features are not linearly separable LDA usually fails. It is possible to use a more generalized approach called Generalized Discriminant Analysis (GDA) (Baudat & Anouar, 2000).

Let the input matrix \mathbf{X} be composed from columns arranged following their class membership. The covariance matrix of the centers of the classes $\Sigma_{\bar{\Phi}(\mathbf{X})}$ is:

$$\Sigma_{\bar{\Phi}(\mathbf{X})} = \frac{1}{M} \sum_{c=1}^C n_c \bar{\Phi}(\mathbf{x}_c) (\bar{\Phi}(\mathbf{x}_c))^T, \quad (40)$$

where n_c is the number of observations belonging to the class c . C is the number of the classes, and $\bar{\Phi}(\mathbf{x}_c) = E\{\Phi(\mathbf{x}_c)\}$. For the covariance matrix $\Sigma_{\Phi(\mathbf{X})}$ the following is valid (Baudat & Anouar, 2000):

$$\Sigma_{\Phi(\mathbf{X})} = \sum_{c=1}^C \sum_{k=1}^{n_c} \Phi(\mathbf{x}_{c,k}) (\Phi(\mathbf{x}_{c,k}))^T, \quad (41)$$

where $\mathbf{x}_{c,k}$ is the k -th observation from the class c . The goal in GDA is maximization of between-class variance and minimization of within-class variance. This is done by the following eigendecomposition:

$$\lambda \Sigma_{\Phi(\mathbf{X})} \mathbf{v} = \Sigma_{\bar{\Phi}(\mathbf{X})} \mathbf{v}, \quad (42)$$

where \mathbf{v} is an eigenvector, and λ is the corresponding eigenvalue. The maximum eigenvalue maximizes the ratio (Baudat & Anouar, 2000):

$$\lambda = \frac{\mathbf{v}^T \Sigma_{\bar{\Phi}(\mathbf{X})} \mathbf{v}}{\mathbf{v}^T \Sigma_{\Phi(\mathbf{X})} \mathbf{v}} \quad (43)$$

This criterion 43 in the space \mathcal{F} is:

$$\lambda = \frac{\mathbf{w}^T \mathbf{K} \mathbf{P} \mathbf{K} \mathbf{w}}{\mathbf{w}^T \mathbf{K} \mathbf{K} \mathbf{w}}, \quad (44)$$

where \mathbf{K} is the kernel matrix, $\mathbf{w} = [w_1, w_2, \dots, w_M]^T$ is a vector from coefficients for which the equation 32 is valid, and $\mathbf{P} [p_{ij}]_{M \times M} = \mathbf{P}_1 \oplus \mathbf{P}_2 \oplus \dots \oplus \mathbf{P}_c \oplus \dots \oplus \mathbf{P}_C$ is a block matrix. The elements of $\mathbf{P}_c [p_{c,ij}]_{n_c \times n_c}$ are:

$$p_{c,ij} = \frac{1}{n_c}, \quad i = 1, \dots, n_c; j = 1, \dots, n_c. \quad (45)$$

The solution of 45 can be found in (Baudat & Anouar, 2000).

3.4 Results and discussion

3.4.1 Datasets for experiments

The ECG signals for the experiments are collected from 28 individuals using own ECG registration hardware. The sampling rate is 512Hz and the resolution is 12bit. The system was trained using subsets from these signals. The testing was performed two weeks later in order to prove the time invariance of the features.

3.4.2 Experimental results

All experiments with kernel versions of PCA and LDA were made using the Gaussian kernel function $k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right)$ with $\sigma^2 = 1$. In Table 2 are presented the achieved values of the accuracy of the ECG personal identification.

Method	Accuracy,%
PCA	78.4
PCA-LDA	79.2
KPCA	91.8
PCA-GDA	95.7
Time domain features (Biel et al., 2001)	95.0

Table 2. Accuracy of the ECG identification

As can be seen the results using holistic features extracted with linear projections in subspaces are relatively poor. The GDA outperforms all approaches but the significant disadvantage of this method is the computation complexity. In addition the maximal dimensionality of the features is limited up to the number of identified individuals minus one. For combining with facial biometric modality we select KPCA approach for feature extraction. Despite its lower performance we prefer it because there is an algorithm, called GreedyKPCA, in which the kernel matrix does not have to be stored.

4. Combining ECG personal identification and face recognition

4.1 Classifier combination approaches

There are different approaches for combining classifiers, depending on their output. If only class labels are available as output, voting schemes can be used for final decision. If *a posteriori* probabilities are available however, different linear combinations rules can be applied. We will follow the former approach.

The strategies for combination utilize the fact that the classifier output reflects its confidence, and not the final decision. The confidence of a single classifier is represented by (Duin, 2002): $P_i(\mathbf{x}) = \text{Prob}(\omega_i | \mathbf{x})$, where ω_i is the i -th class and $i = 1, 2, \dots, M$ is the class number. In the case of multiple classifiers, the confidence need to be defined for each of the classifiers $j = 1, 2, \dots, C$. The output of each classifier can be viewed as a feature vector \mathbf{z}_i . Then, following the Bayesian rule for optimal classification a sample \mathbf{x} is assigned to class ω_i if:

$$P(\omega_i | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_C) = \max_n P(\omega_n | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_C) \quad (46)$$

Using the Bayesian rule we can express 46 by the likelihood functions (Kittler et al., 1998):

$$P(\omega_n | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_C) = \frac{P(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_C | \omega_n)P(\omega_i)}{p(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_C)}, \quad (47)$$

where $p(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_C)$ is the unconditional joint probability density. It can be expressed through the conditional distributions and for this reason it could be used for the calculation of the final decision (Kittler et al., 1998). Also, using the numerator term only allows for various classifier combination rules to be obtained (Kittler et al., 1998):

- **Min rule:** According to this rule a sample \mathbf{x} is assigned to class ω_i if

$$\min_j P(\omega_i | \mathbf{z}_j) = \max_n \min_j P(\omega_n | \mathbf{z}_j), j = 1, 2, \dots, C, n = 1, 2, \dots, M. \quad (48)$$

This rule selects a classifier with least objection against a certain class.

- **Max rule:** In this rule a classifier with a most confidence is selected

$$\max_j P(\omega_i | \mathbf{z}_j) = \max_n \max_j P(\omega_n | \mathbf{z}_j), j = 1, 2, \dots, C, n = 1, 2, \dots, M. \quad (49)$$

This rule selects a classifier with least objection against a certain class.

- **Product rule:** This rule assigns \mathbf{x} to class ω_i if

$$p^{-(C-1)}(\omega_i) \prod_{j=1}^C P(\omega_i | \mathbf{z}_j) = \max_n p^{-(C-1)}(\omega_n) \prod_{j=1}^C P(\omega_n | \mathbf{z}_j), n = 1, 2, \dots, M. \quad (50)$$

- **Sum rule:** This rule can be derived from the product rule and assigns \mathbf{x} to class ω_i if

$$(1 - C)P(\omega_i) + \sum_{j=1}^C P(\omega_i | \mathbf{z}_j) = \max_n \left[(1 - C)P(\omega_n) + \sum_{j=1}^C P(\omega_n | \mathbf{z}_j) \right], n = 1, 2, \dots, M. \quad (51)$$

4.2 Experimental results

We approach the combination of Face and ECG biometric modalities by combining both classifier's output probabilities using the rules specified in the previous section. For ECG identification we use the output probabilities of Radial Basis Neural Network classifier and for Face Recognition framework we use *LIBSVM library* (Chang & Lin, 2001) for calculating the output probabilities of SVM classifier. For both modalities we select 19 persons for identification with 5 samples for training and 5 samples for testing. Hence, the output of both classifiers is a 19 elements feature vector of probabilities with totally 10 samples per person, per modality. In our experiments, we test all rules for classifier combination considered and the results are displayed in Table 3. Also, we compare our work with the best results achieved by *Combining Attributes* in (Israel et al., 2003), .

Combination Rule	Accuracy,%
Max Rule	93.15
Min Rule	99.0
Sum Rule	99.1
Product Rule	99.5
Combining Attributes (Israel et al., 2003)	99.0

Table 3. Accuracy of the ECG identification

Experimental results reveal that combining probabilities output of ECG identification and Face Recognition framework with the *Product Rule* achieved best results.

5. Conclusion

In this work an approach for personal identification based on biometric modality fusion was presented. The presented combination of classifiers is characterized by its high accuracy and it is particularly suitable for precise biometric identification in intelligent video surveillance systems.

6. Acknowledgement

This work was supported by National Ministry of Education and Science of Bulgaria under contract DO02-41/2008 "Human Biometric Identification in Video Surveillance Systems", Ukrainian-Bulgarian R&D joint project References

7. References

- Abate, A. F., Nappi, M., Riccio, D. & Sabatino, G. (2007). 2d and 3d face recognition: A survey, *Pattern Recogn. Lett.* 28: 1885–1906.
- Baudat, G. & Anouar, F. (2000). Generalized discriminant analysis using a kernel approach, *Neural Comput.* 12: 2385–2404.
- Biel, L., Pettersson, O., Philipson, L. & Wide, P. (2001). ECG Analysis – A new approach in Human Identification, *IEEE Transactions on Instrumentation and Measurement* 50(3): 808–812.
- Bishop, C. M. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1 edn, Springer.
- Bolle, R. & Pankanti, S. (1998). *Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society*, Kluwer Academic Publishers, Norwell, MA, USA.
- Boumbarov, O., Velchev, Y. & Sokolov, S. (2009). Ecg personal identification in subspaces using radial basis neural networks, *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 2009. IDAACS 2009. IEEE International Workshop on*, pp. 446–451.
- Cai, D. (2009). *Spectral Regression: A Regression Framework for Efficient Regularized Subspace Learning*, PhD thesis, Department of Computer Science, University of Illinois at Urbana-Champaign.
- Cai, D., He, X. & Han, J. (2007). Spectral regression for efficient regularized subspace learning, *Proc. Int. Conf. Computer Vision (ICCV'07)*.
- Caltech-CV-Group (1999). Faces database.
URL: <http://www.vision.caltech.edu/html-files/archive.html>
- Chang, C.-C. & Lin, C.-J. (2001). *LIBSVM: a library for support vector machines*. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Chang, K. I., Bowyer, K. W. & Flynn, P. J. (2003). Face recognition using 2d and 3d facial data, *ACM Workshop on Multimodal User Authentication*, pp. 25–32.
- Chang, K. I., Bowyer, K. W. & Flynn, P. J. (2005). An evaluation of multimodal 2d+3d face biometrics, *IEEE Trans. Pattern Anal. Mach. Intell.* 27: 619–624.
- Chung, F. (1997). *Spectral Graph Theory*, American Mathematical Society.
- Duin, R. P. W. (2002). The combining classifier: to train or not to train?, *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, Vol. 2, pp. 765–770.
- Franc, V. & Hlavac, V. (2004). *Statistical Pattern Recognition Toolbox for Matlab, Prague, Czech: Center for Machine Perception, Czech Technical University*.
URL: <http://cmp.felk.cvut.cz/cmp/software/stprtool/index.html>

- Freund, Y. & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting, *J. Comput. Syst. Sci.* 55: 119–139.
- Hotelling, H. (1931). The generalization of student's ratio, *The Annals of Mathematical Statistics* (3): 360–378.
- Hughes, N. P., Peter, N. & Philosophy, H. D. (2004). Markov models for automated ecg interval analysis, in *Proceedings NIPS 16*, MIT Press, pp. 611–618.
- Israel, S. A., Irvine, J. M., Cheng, A., Wiederhold, M. D. & Wiederhold, B. K. (2005). Ecg to identify individuals, *Pattern Recogn.* 38: 133–142.
- Israel, S. A., Scruggs, W. T., Worek, W. J. & Irvine, J. M. (2003). Fusing face and ecg for personal identification, *Applied Image Pattern Recognition Workshop*, 0: 226.
- Kittler, J., Society, I. C., Hatef, M., Duin, R. P. W. & Matas, J. (1998). On combining classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20: 226–239.
- Kurylyak, Y., Paliy, I., Sachenko, A., Chohra, A. & Madani, K. (2009). Face detection on grayscale and color images using combined cascade of classifiers, *Computing* 8: 61–71.
- Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998). Gradient-based learning applied to document recognition, *Proceedings of the IEEE* 86(11): 2278–2324.
- Malmivuo, J. & Plonsey, R. (1995). *Bioelectromagnetism : Principles and Applications of Bioelectric and Biomagnetic Fields*, 1 edn, Oxford University Press, USA.
URL: <http://www.worldcat.org/isbn/0195058232>
- Otsu, N. (1979). A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man and Cybernetics* 9(1): 62–66.
- Plataniotis, K., Hatzinakos, D. & Lee, J. (2006). Ecg biometric recognition without fiducial detection, *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the*, pp. 1–6.
- Sanei, S. & Chambers, J. A. (2007). *EEG Signal Processing*, Wiley-Interscience.
URL: <http://www.worldcat.org/isbn/0470025816>
- Schölkopf, B., Smola, A. & Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem, *Neural Comput.* 10: 1299–1319.
- Simard, P. Y., Steinkraus, D. & Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis, *Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2, ICDAR '03*, IEEE Computer Society, pp. 958–.
- Theodoridis, S. & Koutroumbas, K. (2006). *Pattern Recognition, Third Edition*, Academic Press.
URL: <http://www.worldcat.org/isbn/0123695317>
- Vapnik, V. N. (1998). *Statistical Learning Theory*, Wiley-Interscience.
URL: <http://www.worldcat.org/isbn/0471030031>
- Velchev, Y. (2010). Using morphological filtering and histogram analysis, *Computer and Communications Engineering* 4(1): 51–55.
- Viola, P. & Jones, M. J. (2004). Robust real-time face detection, *Int. J. Comput. Vision* 57: 137–154.
- Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q. & Lin, S. (2007). Graph embedding and extensions: A general framework for dimensionality reduction, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29(1): 40–51.