Control and Cybernetics

vol. **41** (2012) No. 3

Acces control system using face image*

by

Janusz Bobulski

Czestochowa University of Technology
Institute of Computer and Information Science
Dabrowskiego 73, 42-200 Czestochowa, Poland
januszb@icis.pcz.pl

Abstract: Ensuring safety requires the use of access control systems. Traditional systems typically use proximity cards. Modern systems use biometrics to identify the user. Using biological characteristics for identification ensures a high degree of safety. In addition, biological characteristics cannot be neither lost nor stolen. This paper presents proposals for the access control system using face image. The system operates in real time using camera image.

Keywords: face localization, face detection, face recognition, biometrics, access control

1. Introduction

Access control is the most often implemented security system. It is difficult to imagine a modern company or institution not using this kind of system. Such system makes unauthorized persons have limited access to some compartments or zones. Access control is of great significance when it comes to protecting a building, persons, resources and information. Access control systems usually work on the basis of proximity card, which has its individual serial number encoded in its memory. When the card is held close to a reader unit, the device decodes the number used for the user's identification. Unfortunately, this kind of systems does not control who uses the card: an owner or an unauthorized person. Additionally the card may be lost or stolen. Alternative systems are the ones using biometrics, which do not have the above mentioned disadvantages. The system uses individual and unique biological features of a person for identification. The most popular features are: fingerprint, iris, voice and face image. Systems basing on face image have one additional attribute, which the others do not. The user may not be aware of the fact that his/her identity is being verified. The pattern may be taken through a hidden camera, whereas systems using fingerprint or iris require the person being close to the reader unit. Such feature is most frequently seen as a deficiency of access control systems, as users associate it with being treated like a criminal.

*Submitted: February 2011; Accepted: August 2012

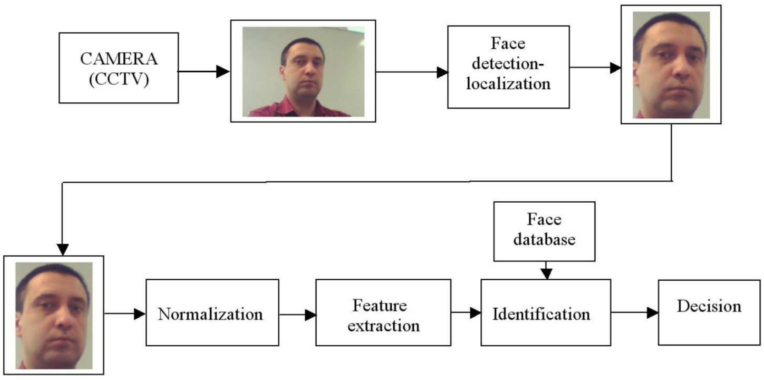

Figure 1. Scheme of real-time face recognition system

2. The proposed system

The control system has to work in the real time to perform the above mentioned tasks. In the case of a person identification based on facial image, it becomes a difficult task due to its computational complexity. The proposed access control system consists of the following units: image obtainment, face detection and localization, face normalization, feature extraction and identification. Scheme of the system is shown in Fig.1. Hybrid method of face localization using skin detection and template matching is applied in the first stage of the system. When face is localized, then it is normalized. Next, wavelet transform is used to extract features. Hidden Markov models are used for training and testing procedure during identification process and the final decision is based on maximum likelihood.

3. Face detection

Face detection is the first part of processing path. The success of identification (recognition) depends on effectiveness of face detection and localization. There are many methods of face detection. The most popular methods are the ones based on skin colour detection. They use some colour space such as RGB, HSV, YCbCr and others (see Nabiyev and Gnay, 2008, or Vezhnevents et al., 2010, or Kukharev and Kuzminski, 2003). Disadvantages of these techniques are: many false positive errors and sensitivity to change of lighting condition and type of lighting (bulb, fluorescent, sun). The second group of face detection methods consists of methods using the template matching. The idea of these techniques is to compare an input image with the pattern including a face (see Brunelli, 2009). They have good recognition rate, but they are computationally expensive, because they need the whole image to be analysed. The methods using features constitute the third group of techniques of face detection. They may use Eigenface (PCA/KLT), Hidden Markov Models, Support Vector Machines or statistics. They are very effective but also complicated and computationally expensive.

3.1. Skin colour detection

The face localization over colour image uses technique named skin colour detection. This consists in quantization of colourspaces, segmentation of image and next, separation of skin regions. It needs to verify each region whether it is a face or not. The most popular colourspaces are: RGB, HSV, YCbCr.

RGB

RGB is a colourspace originated from display devices, describing colour as a combination of three coloured rays: Red, Green and Blue. It is one of the most widely used colour spacees for processing and digital image data storing (see Vezhnevents et al., 2010). The simple segmentation on RGB colourspace may be used (1) (see Kubanek, 2006). Fig. 2 presents the result of application of this technique,

$$\frac{R}{B} - \frac{G}{B} > threshold \quad R, G, B = 0..255. \tag{1}$$

HSV

Hue–saturation–based colour spaces were introduced when the user needs to specify colour properties numerically. Spaces describe colour with intuitive values, based on the artistic idea of tint, saturation and tone. Hue defines the dominant colour (such as red, green, purple and yellow) of an area, saturation measures the colourfulness of an area in proportion to its brightness. The value relates to the colour luminance (see Vezhnevents et al., 2010). It may be used for skin detection (2) (see Kukharev and Kuzminski, 2003). The result of application of this technique is shown in Fig. 3

$$\begin{gathered} RGB \to HSV, \ H < 0.15 \\ or \ H > 0.95, \ S < 0.68, \ V > 0.3, \ H, S, V = [0, 1]. \end{gathered} \tag{2}$$

Values of *H, S, V* for skin detection are selected experimentally.

YCbCr

YCrCb is an encoded non-linear RGB signal, commonly used by European television. Colour is represented by Y, which is luminance, computed from non-linear RGB, constructed as a weighted sum of the RGB values, and two colour difference values of chrominance Cr and Cb that are formed from components

Figure 2. RGB mask (1) for skin detection

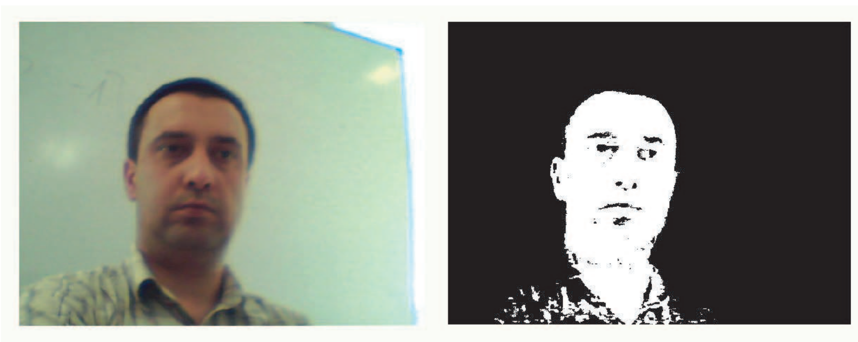

Figure 3. HSV mask for skin detection

of RGB (see Vezhnevents et al., 2010, or Kukharev and Kuzminski, 2003). The result of application of this technique (3) is presented in Fig.4

$$Y > 80, \quad 85 < Cb < 135, \quad 135 < Cr < 180, \quad Y, Cb, Cr = 0..255. \tag{3}$$

3.2. Template matching

The template matching technique is used to classify an object, by comparing a part of an image with another one and may be used to recognize a similar object. The template matching method for face detection uses the pattern of face (see Brunelli, 2009, or Ahlvers et al., 2005). Face pattern is being compared with the whole input image from the top to the bottom, and from the left to the right. It is computationally very expensive. Additionally, these types of face detection methods are characterized by frequent non-face errors.

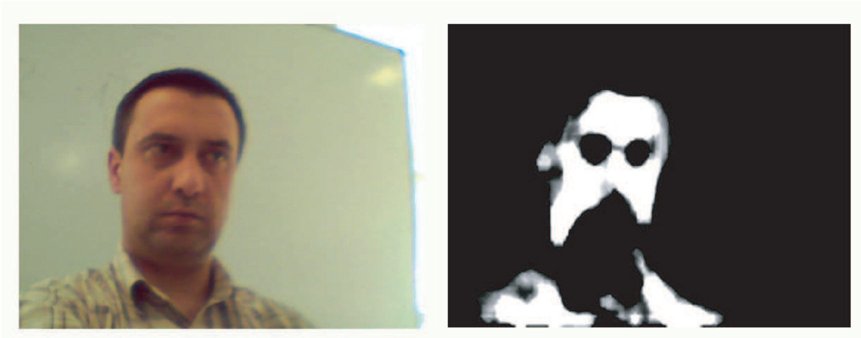

Figure 4. YCbCr mask for skin detection

3.3. The proposed method of face localization

The above presented face detection methods have both advantages and disadvantages. My own detection method is presented below. This method uses skin detection technique to select potential areas containing a face. RGB colour space is used in this stage. Next, template matching technique is applied to verify each region if it is face or not. When we compare this method with the other ones we find out that it uses the image of eyes as a pattern (Fig. 5) and not the whole face as used in most methods. We have the following algorithm:

1. Start
2. Take a frame (image) from video stream.
3. Make a quantization to HSV colour space.
4. Detect skin
5. Make a mask.
6. Select skin areas.
7. For i=1 to *numbers_of_skin_areas*
   (a) Take size of area.
   (b) Scale pattern of eyes adequately to the size of skin area.
   (c) Compare pattern with area: if *distance* < *threshold* than save coordinates of face.
8. Go to 2.

4. User identification

The user identification process works on the basis of the frontal facial image, in which the fusion of Wavelet Transformation (WT) and Hidden Markov Models (HMM) are used for each of three face parts independently (eyes, nose and mouth). The likelihood of generating observation is computed for each part of face. The decision is made on the basis of the maximum likelihood sum. The most popular method of face identification is Principal Component Analysis

Figure 5. Samples of eye pattern

(PCA) (see Kukharev and Kuzminski, 2003, or Kirby and Sirovich, 1990). Other popular methods use Wavelet Transform (see Chien and Wu, 2002), Neural Networks (see Bae and Kim, 2008), SVM (see Yand et al., 2008) or Hidden Markov Models (see Samaria and Young, 1994). Analysis of the existing solutions revealed their defects, which caused their weak effectiveness. The disadvantages of these methods are as follows:

- In case of new user registration, process of learning and adding his/her facial image to a database requires repeated learning of the whole system.
- They work with whole face.
- They are computationally very expensive.

The proposed method is a combination of two mathematical tools, Wavelet Transform (WT) and Hidden Markov Model (HMM). Both were mainly used for speech recognition. Here, WT is used for feature extraction, and HMM for identification. This system works in two modes, learning and testing. These modes differ from each other. The algorithm of this method consists of four main parts:

1. Pre–processing: normalization and face division into three parts
2. Feature extraction: WT of the face image
3. Training: generating and learning HMM for each part of the face
   Testing: testing models from the database
4. Training: saving learned models of face to database
   Testing: making a decision - maximum likelihood of the model.

4.1. Pre–processing

The input image is made up of 300 x 400 pixels. The normalization consists of fixing the eye centres, and then face scaling, so that the distance between eye centres equals 120 pixels* and only face area is drawn out. Final image has the same size of 300 x 400 pixels. The second part of this process is dividing the normalized face into three parts: the area of eyes, nose, and mouth (Fig.6). Each of these regions has pixel size of 300 x 140. There are 30 pixels of overlap

*Face normalization is made according to the international norm about face image processing in biometrics - ISO/IEC JTC 1/SC 37 Biometric Data Interchange Formats Part 5: Face Image Data

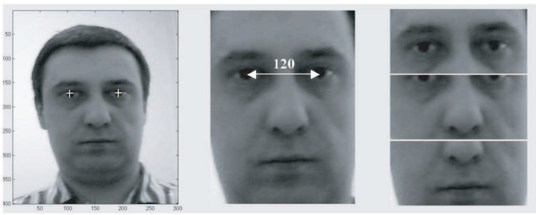

Figure 6. Pre-processing of the face image

between regions.

4.2. Feature extraction

WT is used for feature extraction. Using 2D WT (Fig. 7), the face image is decomposed into four sub–images via the high–pass and low–pass filtering. The image is decomposed along column direction into sub images to high-pass frequency band H and low-pass frequency band L. Assuming that the input image is a matrix of m x n pixels, the resulting sub images become $m/2$ x n matrices. During the second step the images H and L are decomposed along row vector direction and respectively produce the high and low frequency band HH and HL for H, and LH and LL for L. The four output images become the matrices of $m/2$ x $n/2$ pixels. Low frequency sub image LL (A_1) possesses high energy and is a smaller copy of the original image (A_0). The remaining sub–images LH, HL, and HH, respectively, extract the changing components in horizontal (D_{11}), vertical (D_{12}), and diagonal (D_{13}) direction (see Garcia et al., 2000).

Wavelet Transform of the second level (Fig. 8) is used for feature extraction in the proposed technique. After the first level wavelet decomposition the output images become input images of the second level decomposition.

The results of two-level 2D WT are coded so that they may be applied in HMM (Fig. 10). One of the simplest methods of reduction and information coding is calculating standard deviation. Each part of the face is transformed separately by discrete wavelet transform (Fig. 9). The features vector is formed separately for each face part. Standard deviation is calculated for each column of the global image, which is the result of the second level wavelet decomposition. Effect of this computation is quantized according to the number of observation in HMM. The size of feature vectors is 300

$$S(X) = \sqrt{\frac{1}{n}\sum_{n=1}^{n}(X_i - \overline{X})^2} \tag{4}$$

n- number of image columns
i- index of image column

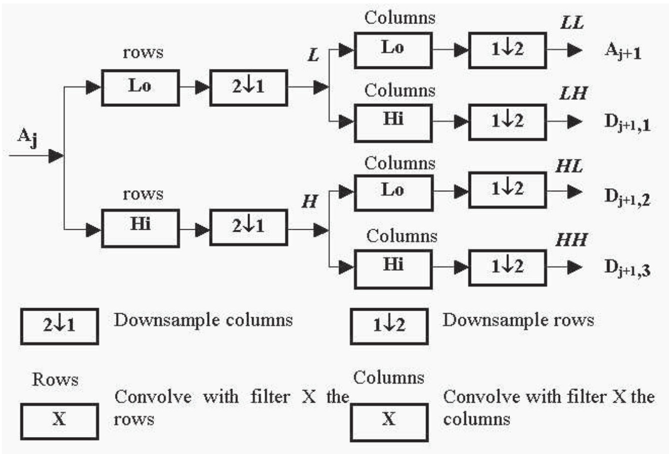

Figure 7. Scheme of one-level two-dimensional wavelet transform (see Misiti et al., 2002)

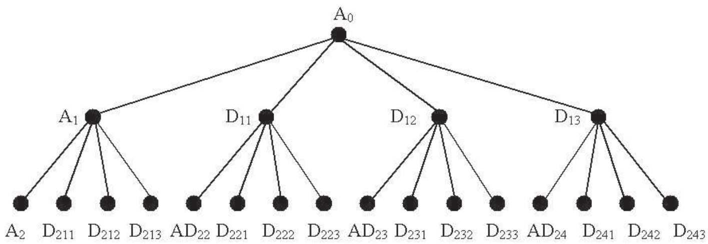

Figure 8. The wavelet decomposition tree

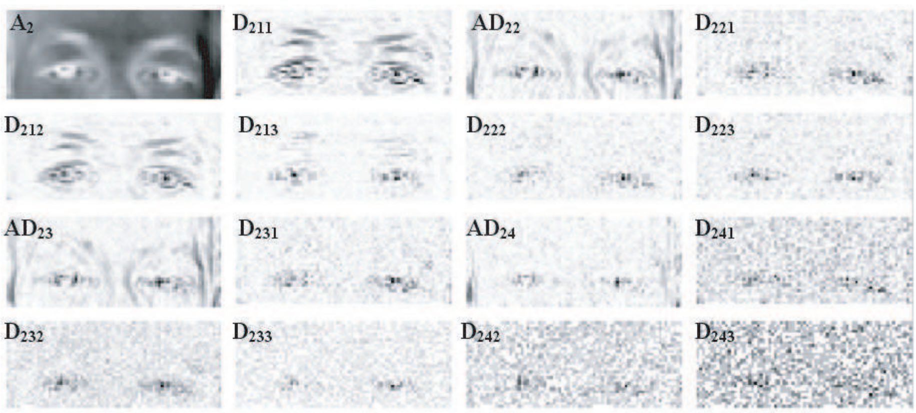

Figure 9. Example of level 2 of the wavelet decomposition of eyes area image

X- image column
$\overline{X}$ - mean value.

Feature vector structure of face parts.

$$W1 = S(MW1); W2 = S(MW2); W3 = S(MW3) \quad (5)$$

MW- wavelet coefficients matrix.

The bank filter selection is an important aspect of this transformation. It guarantees a good recognition rate (see Kozuani and Sammut, 1997). The experiment results related to wavelet function selection are shown in Table 1.

Table. 1. Experimental results of wavelet function selection

Wavelet function	Number of correct identified faces	Number of false identified faces	Recognition rate [%]
db1	18	2	90
db2	14	6	70
db3	13	7	65
Sym2	16	4	80
Coif1	12	8	60
bior1.1	15	5	75

4.3. Training system

HMM is used for the identification process. HMM is a double stochastic process with underlying stochastic process that is not observable (hidden), but can be observed through another set of stochastic processes that produce a sequence of observations. Let $O = \{O_1, \ldots, O_T\}$ be the sequence of observations of feature vectors, where T is the total number of feature vectors in the sequence. The statistical parameters of the model may be described as follows (see Rabiner, 1989):

- The number of states of the model N
- The transition probabilities of the underlying Markov chain, $A = \{a_{ij}\}, 1 \leq i, j \leq N$, where a_{ij} is the probability of transition from state i to state j
- The observation probabilities, $B = \{b_j(O_T)\}, 1 \leq j \leq N, 1 \leq t \leq T$ which represents the probability of the t^{th} observation conditioned on the j^{th} state
- The initial probability vector, $\Pi = \{\pi_i\}, 1 \leq i \leq N$.

Hence, the HMM requires three probability measures to be defined - A, B, Π. Notation $\lambda = (A, B, \Pi)$ is often used to indicate the set of parameters of the model. In the proposed method, one model is made for each face part. The parameters of the model are generated at random at the beginning. Then, they are estimated with the Baum-Welch algorithm, which is based on the forward-backward algorithm. The forward algorithm calculates the coefficient $\alpha_t(i)$ (probability of observing the partial sequence $(o_1, \ldots, o_t)$ such that the

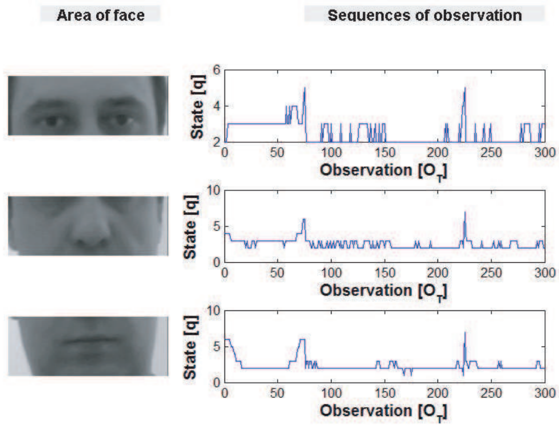

Figure 10. Part of face and corresponding them sequences of observation

state q_t is i). The backward algorithm calculates the coefficient $\beta_t(i)$ (probability of observing the partial sequence $(o_{t+1}, \ldots, o_T)$ such that state q_t is i). The Baum-Welch algorithm can be described as follows (see Kanungo, 1999):

1. Let initial model be λ_0
2. Compute new λ based on λ_0 and observation O
3. If $log(P(O|\lambda)log(P(O)|\lambda_0) < DELTA$ stop
4. Else set $\lambda_0 \leftarrow \lambda$ and go to step 2.

The parameters of the new model λ, based on λ_0 and observation O, are estimated from the equation of the Baum-Welch algorithm (see Kanungo, 1999), and then are recorded in the database.

4.4. Testing system

The testing process consists of computing the probability of observation generated by models saved in database and choosing the model with maximum likelihood. In the proposed method, probabilities are calculated separately for each of the three models representing face parts, and then they are summed. The face, for which the sum of probability is maximal, is chosen as the correct face. The probability of generating sequences of observations is computed from the equations 6-9 (see Kanungo, 1999).

$$P(O|\lambda) = \sum_q P(O|q, \lambda)P(q|\lambda) \tag{6}$$

$$P(O|q, \lambda) = \prod_{i=1}^{T} P(o_t|q_t, \lambda) = b_{q_1}(o_1)b_{q_2}(o_2)\ldots b_{q_T}(o_T) \tag{7}$$

$$P(q|\lambda) = \pi_{q_1} a_{q_1q_2} a_{q_2q_3} \ldots a_{q_{T-1}q_T} \tag{8}$$

$$PF = \sum_{i=1}^{3} P(O_i|\lambda_i). \quad (9)$$

5. Experimenting

The system has been tested in real environment and obtained 75% recognition rate, which is the result of error sum from two modules: face detection and recognition.
The face detection module is sensitive to face rotation and different lighting conditions and types, giving face detection rate of 83%. It was tested on 64 persons moving in front of camera. The observed faces were in different poses and light conditions. Positive detection was achieved for 48 persons. The face recognition module is sensitive to face rotation and that is why its face recognition rate is 92%. This module was tested on two face databases BioID and FaDaB (own database). Characteristics of this database are shown in Table 2. Two images were used for learning and one image for testing.

Table 2. Face database characteristics

Parameter	BioID	FaDab
Number of images	1521	450
Number of persons	24	150
Colour palette	Grey scale	Grey scale
Depth of colour	256	256
Size of image [high x width in pixels]	450 x 345	400 x 300
Resolution [dpi]	70	70
File Format	*.bmp	*.bmp
Emotion	No	No

Table 3. Recognition experiment results

Face database	Number of face parts	Number of persons	Number of persons false recognition	Error rate [%]
BioID	3	24	3	12.51
BioID	1 (eyes)	24	2	8,33
FaDab	3	150	15	10.00
FaDab	1 (eyes)	150	12	8.00

This system gives acceptable results compared to other methods (Tables 3–5). The proposed system distinguishes from the other methods with the following advantages:

- fusion of HMM and WT gives better results than HMM alone, see Samaria and Young (1994);
- there is no problem with class definition and separation like in Eigenvector/PCA (see Kukharev and Kuzminsky, 2003);

- method is less computationally expensive than NN (see Bae and Kim, 2005) and SVM (see Yang et al., 2008);
- method does not need so much learning data as NN.

Table 4. Comparison of face detection methods

Method	Detection rate [%]
Eigenvector	92
SVM	75
Neural Network	81
Template matching	80
Proposed skin&eye	83

Table 5. Comparison of face recognition methods

Method	Recognition rate [%]
Eigenvector	94
HMM	84
Neural Network	91
SVM	96
Proposed HMM&WT	92

6. Conclusion

The article presents control system of user access based on face recognition method working in real-time. Experimental recognition rate of the whole system is 75%. Error rate (25%) is the result of superposition of errors of detection face module (17%) and recognition module (8%). The whole system is sensitive to face rotation. Future work will be concentrated on eliminating this problem by detection and rotation of face along the X axis, and adding other poses of the face with the aim of eliminating face rotation along axes X and Z. The proposed method of user identification is characterized by usage of three areas of the face for identifying and creating one independent HMM for each of them (which makes it possible to use them separately or together). This procedure allows for lowering computational complexity and can obtain a good recognition rate. On the basis of experimental studies it was stated that the eyes area contains the most useful information for person identification, and could be successfully applied in specific identification methods. Additionally, an advantage of the method is that in case of a new user registration, the process of learning and adding his/her facial image to a database does not require repeated learning of the whole system.

References

AHLVERS U., ZÖLZER U., RAJAGOPALAN R. (2005) Model–free Face Detection and Head Tracking with Morphological Hole Mapping. *EUSIPCO'05*, Antalya, Turkey.

BAE H. and KIM S. (2005) Real-time face detection and recognition using hybrid–information extracted from face space and facial features. *Image and Vision Computing* 23, 1181-1191.

BRUNELLI R. (2009) *Template Matching Techniques in Computer Vision: Theory and Practice*, Wiley.

CHEN S., ZHANG T., ZHANG Ch., CHENG Y. (2010) A real–time face detection and recognition system for a mobile robot in a complex background. *Artif Life Robotics* 15, 439-443.

CHIEN J.T. WU Ch.Ch. (2002) Discriminant Waveletface and Nearest Feature Classifiers for Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24** (12), 1644-1649.

GARCIA C., ZIKO, G., TZIRITAS G. (2000) Wavelet pocket analysis for face recognition. *Image and Vision Computing* 18, 289-297.

HJELMAS E., LOW B.K. (2001) Face Detection: A Survey. *Computer Vision and Image Understanding* 83, 236–274.

KANUNGO T. (1999) *Hidden Markov Model Tutorial.* www.kanungo.com/ software/hmmtut.pdf.

KIRBY M., SIROVICH L. (1990) Application of the Karhunen–Loeve Procedure for the Characterization of Human Faces. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **12** (1), 103–108.

KOZUANI A.Z., HE F., SAMMUT K. (1997) Wavelet Packet Face Representation and Recognition. *Proc. IEEE Conf. Systems, Man, and Cybernetics*, NYC, 1614–1619.

KUBANEK M. (2006) Method of speech recognition and speaker identification with use audio–visual of polish speech and hidden markov models. In: K. Saeed, J. Peja, R. Mosdorf, eds., *Biometrics, Computer Security Systems and Artificial Intelligence Applications.* Springer Science + Business Media, 45–55.

KUKHAREV G., KUZMINSKY A. (2003) *Techniki Biometryczne* In: Metody Rozpoznawania Twarzy. Politechnika Szczecińska, Szczecin.

MISITI M., MISITI Y., OPPENHEIM G., POGGI J.-M. (2002) *Wavelet Toolbox Users Guide.* The MathWorks.

NABIYEV V., GAŚNAY A. (2008) Towards A Biometric Purpose Image Filter According To Skin Detection. *Problems of Cybernetics and Informatics Conference*, Baku, Azerbaijan. www.pci2008.science.az/2/17.pdf

RABINER L. R. (1989) A tutorial on hidden Markov models and selected application in speech recognition. *Proc. IEEE* 77, 257–285.

SAMARIA F, YOUNG S. (1994) HMM–based Architecture for Face Identification. *Image and Vision Computing*, **12**, 8, 537–583.

VEZHNEVENTS V., SAZONOV V., ANDREEVA A.(2010) *A Survey on Pixel Based Skin Color Detection Techniques*, http://academic.aua.am/Skhachat/Public/Papers

YANG J., LING X., ZHU Y., ZHENG Z. (2008) A face detection and recognition system in color image series. *Mathematics and Computers in Simulation* 77, 531-539.