



FACE RECOGNITION SYSTEM FOR MACHINE READABLE TRAVEL DOCUMENTS

Sergey Tulyakov ¹⁾, Rauf Kh. Sadykhov ²⁾

Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus
sergei.tulyakov@gmail.com, rsadykhov@bsuir.by

Abstract: *This paper presents an upright frontal face recognition system, aimed to recognize faces on machine readable travel documents (MRTD). The system is able to handle large image databases with high processing speed and low detection and identification errors. In order to achieve high accuracy eyes are detected in the most probable regions, which narrows search area and therefore reduces computation time. Recognition is performed with the use of eigenface approach. The paper introduces eigenface basis ranking measure, which is helpful in challenging task of creating the basis for recognition purposes. To speed up identification process we split the database into males and females using high-performance AdaBoost classifier. At the end of the paper the results of the tests in speed and accuracy are given.*

Keywords: *face detection, face recognition, eigenface basis ranking measure, gender determination.*

1. INTRODUCTION

During two last decades the technologies of face detection and recognition have attracted considerable research interest. These technologies are used in biometric systems, border control systems, video surveillance, human-computer interface, access control systems, face expression recognition, content based image retrieval.

The human ability to recognize faces is remarkable. We can easily recognize faces that have seen only several times. Such skills are quite robust to time, face expression, pose, and other distractions such as presence or absence of beard, mustache, glasses, traces of injuries and surgical operations.

However, the tasks of face detection and recognition are serious problems for automation due to many factors such as orientation, imaging conditions, presence or absence of structural components such as beards, mustache, glasses [1].

Methods to detect faces can be divided into four main groups: knowledge-based methods, feature invariant approaches, template matching methods and appearance-based methods [1]. Knowledge-based methods use expert knowledge to encode typical face. Feature invariant approaches are aimed to find facial features even when imaging conditions (pose, viewpoint or illumination) vary. Template matching methods store several standard patterns to describe a whole face or facial features separately.

Nowadays the appearance-based methods gained

most popularity due to their hit rate and speed [1]. These methods don't use a priori knowledge in the data that is present on the image. Instead, they try to extract different variation modes of the database and provide a set of subclasses which represent them best. Appearance-based methods are based on scanning the input image on different scales with fixed window size in order to find faces. Each window then is given as an input to some classifier, trained to separate two classes of objects (face/non-face).

There are many models that can be used as a classifier. To separate faces and non-faces Osuna et al. [2] and later Romdhani et al. [3] use approach based on support vector machines. Schneiderman propose to apply a statistical method to the problem of face detection [4]. Rowley et al. [5] present a face detection system based on artificial neural network. This approach has become an early standard in face detection.

Later Viola and Jones present a much faster detector than any of their contemporaries [6]. They use rectangular features instead of using pixels directly. The reason for this is that feature-based systems operate much faster. In order to compute rectangular features very fast they introduce an intermediate representation for the image called integral image. This integral image is used as an input for the cascade of weak classifiers. As a learning algorithm they use AdaBoost learning [7].

Our approach is based on cascade of weak

classifiers. In order to increase detection speed and hit rate we introduce some heuristics that will be discussed in the next sections. Recognition stage is done using eigenface approach [8].

To speed up the identification process of a particular individual we split the database into groups according to gender information. There are several known algorithms capable of determining gender with high performance and accuracy. Moghaddam and Yang in [10] use support vector machines to classify gender with the use of thumbnail images. In [11] AdaBoost classifier is applied to features proposed by Viola and Jones in [6] to detect faces. We have chosen the approach proposed by Baluja and Rowley in [12] for simplicity and efficiency: gender can be classified with the comparison of a small number of pixels in the image. The authors have run various experiments to prove that their approach outgoes competitors in performance and accuracy.

2. SYSTEM OVERVIEW

Our system works with Machine Readable Travel Documents (MRTD). MRTD is an official document, conforming to the specifications contained in Doc 9303 [13].

MRTD portraits must conform the quality requirements contained in Doc 9303 (e.g. the portrait shall be color neutral showing the applicant with the eyes open and clearly visible; the portrait shall show the eyes clearly with no light reflection off the glasses and no tinted lenses).

The proposed face recognition program consists of several stages (see Fig. 1).

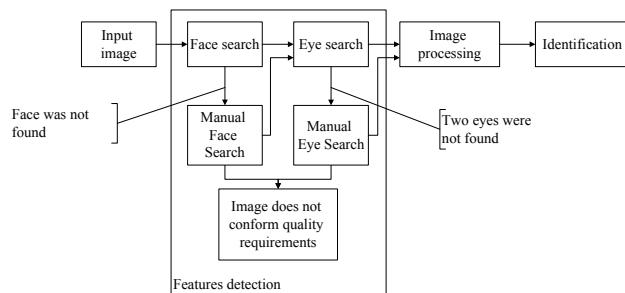


Fig. 1 – Structure of the proposed computer face recognition system

After input image is loaded, system starts feature detection stage. This stage consists of face search and eye search. If system failed to find face or eyes on image it prompts the user to mark them manually. If the user didn't manage to find features, then the image either does not conform to quality requirements or is not a portrait at all. When this stage is successfully completed, system moves to the next stage, where the image processing starts. This stage ends up with storing the image prepared for

identification in the database. Identification stage uses known images to determine the most similar individual to the query one.

3. FEATURES DETECTION

The aim of this stage is to find candidates to be faces and to find eyes on them. To detect features the proposed system uses cascade classifier introduced by Viola et al. [6] along with extended set of Haar-like features proposed by Lienhart et. al. [14].

As mentioned before appearance-based methods scan input image at some scale levels with a fixed-size window. So, if region represents a face, it is found several times. If we know exactly that there is only one face in the image, we should choose region with maximum number of neighbors.

Image B in Fig. 2 contains two face-candidates. Left one was found 15 times, whereas right one was found 178 times (white label in the upper-left corner of the rectangle). Thus, as we know, that there is only one face on MRTD, we should consider right face-candidate to be a face. Alternatively, if we use “find the biggest face” strategy, we would also be able to reject false hit.

We have run experiments in order to determine which approach is better and “the biggest face approach” showed six times higher speed, while accuracy remains the same.

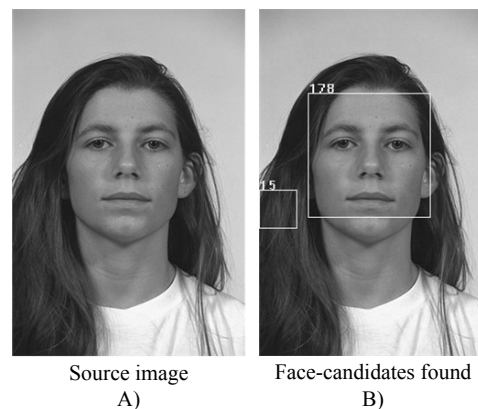


Fig. 2 – Face-candidates found

In order to increase speed of eye search we narrow the region of interest. We have computed regions on image that most likely will contain eyes. These regions are determined using masks that are applied to face area. This idea helps significantly increase detection speed. If eyes were found inside these regions the system proceeds to next step. Here we can also apply “the biggest eye” approach. But our experiments showed that in terms of hit rate it is better to search for eye regions that have most neighbours.

If there were no eyes in predicted regions found this can mean the following: eyes are closed or incorrectly covered with glasses, image has low

quality or image is rotated. The correction of image

First, each image is rotated (see Fig. 3 A). Fig. 4

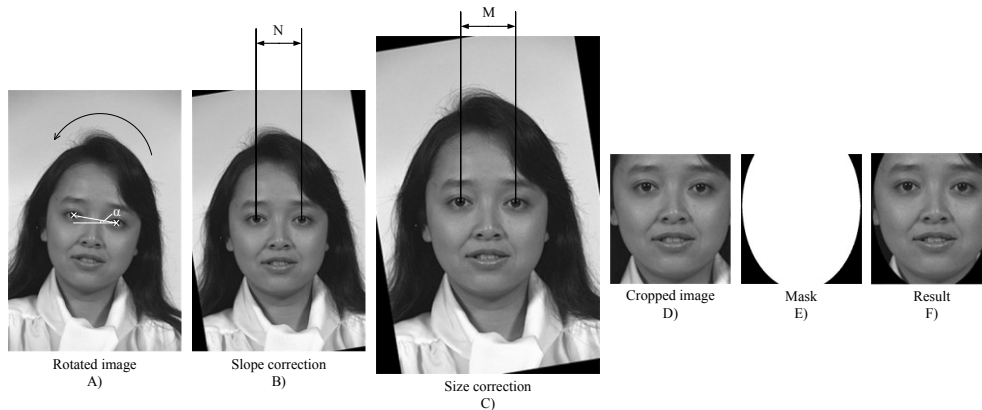


Fig. 3 – The sequence of image transformations

slope is discussed in the next section.

Due to the fact that appearance-based methods scan face at different scales, the scale parameter influences the detection speed and accuracy. By saying accuracy we mean the precise detection of face and eye point: the closer the centre of the eye rectangle is to the pupil of the eye the greater is the accuracy. Big scale parameter yields low accuracy and high speed, low scale parameter – high accuracy and low speed. However, only a small number of images require low scale parameter, most of the images are rapidly and accurately recognized even with high scale parameter.

This tells us to use variable scale parameter for rapid detection on easy images, while allowing accurate detection on complicated ones. We set separate boundaries for the value of scale parameter for face detection and eye detection. Detection starts with the highest value, which is subsequently lowered to the minimum value, only if the object was not found. This approach dramatically improved accuracy. Moreover, images being treated as not containing faces with fixed parameter were successfully recognized with scale parameter. These results are shown in the Table 1.

Table 1. Comparison of Variable and Fixed Scale Parameter

	Fixed			Variable
	1,4	1,25	1,1	
True positive rate	0,55	0,72	0,91	0,98
Speed, image/seconds	8,45	8,13	5,11	7,52

4. IMAGE PROCESSING

The goal of this step is to adjust images so that: imaginary line drawn between eyes is parallel to the top edge of the image, distance between eyes is set to a certain value and image is cropped and masked.

illustrates three images that were initially rotated by 23 (A), 24 (B) and 25 (C) degrees and corrected by the system. Imaginary line drawn between eyes on images B and C is not parallel to top edge of the image. So, we can expect our system to be able to correct image rotations up to approximately 23 degrees. Since MRTD are not likely to be rotated by more than 23 degrees the ability of the system to correct slope is sufficient.

Second, the distance between eyes is measured, and every image is resized to set this parameter to a certain value. This is done to ensure that all images have the same size in pixels (Fig. 3 B, C).

Let us consider that all images in our database have M pixels between the eyes. So, in order to be able to recognize new face we need to correct the size of the new image using the factor M/N , when new face has N pixels between eyes.

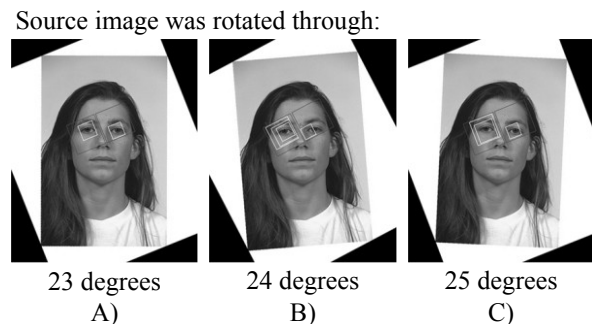


Fig. 4 – Maximum slope correction

Next, each image is cropped to reduce image size. Image C on the Fig. 3 will be stored in the database in order to help the user of the system identify person manually. Image F is used to identify person automatically.

At this step we warp image using only eye regions. If we consider resizing image using both x and y axis, we change the proportions of the face and that will probably decrease recognition abilities of the system. After this stage is complete we have a set of extracted faces.

5. IDENTIFICATION

The goal of this step is to find individuals which are the most similar to the input individual. In our work we use eigenfaces approach [8]. The reason is that this approach is able to handle large amounts of data. Face image is considered to be a two-dimensional N by M array. An image can also be treated as a vector of dimension $N \cdot M \times 1$. Thus, each image is considered to be a point in the $N \cdot M$ -dimensional subspace. Face images, being similar in overall configuration, will not be randomly distributed in this huge image space. In order to decrease dimensionality principal component analysis (PCA) is used. This approach consists of the following steps: (i) compute the average face vector Ψ , (ii) obtain a set of face vectors Φ centered at expectation Ψ , (iii) compute the covariance matrix and find its K eigenvectors with the biggest eigenvalues and (iv) project each image onto new subspace to obtain coordinates of each image in this subspace.

Average face can be computed as follows:

$$\Psi = \frac{1}{n} \sum_{i=1}^n F_i, \quad (1)$$

where F is training set of face vectors, n – number of faces in training set. Then, centered at expectation set of face vectors is $\Phi_i = F_i - \Psi$. Average face Ψ is computed using faces of people of different race, gender, age, color, with and without glasses, beards and mustache.

Our experiments showed that if we increase n , average face remains almost the same. The Euclidean distance between mean faces based on 100 and 200 images is 4.01; 200 and 300 images – 3.3; 300 and 400 images – 0.66; 400 and 500 images – 0.58.

Covariance matrix is determined as follows:

$$C = AA^T, \quad (2)$$

where $A = (\Phi_1 \Phi_2 \dots \Phi_n)$ is $N \cdot M \times n$ matrix, and C is $(N \cdot M)^2$ matrix. Matrix C is too large to compute its eigenvectors v_i . Turk et al. [8] propose to compute v_i as follows:

$$v_i = Au_i, \quad (3)$$

where u_i are eigenvectors of $n \times n$ matrix $A^T A$. Now each face can be represented as a linear combination of first K eigenvectors as follows:

$$\Phi_i = \sum_{j=1}^K w_j v_j, \quad (4)$$

where $w_j = v_j^T \Phi_i$. Fig. 5 shows image

decomposition to mean face and eigenfaces with biggest weights.

In our work we use Euclidean distance to classify an image:

$$\varepsilon_k^2 = (\Omega - \Omega_k)^2, \quad (5)$$

where Ω_k is a vector describing k th face class which minimizes ε_k^2 . A face is classified as belonging to class k when the minimum ε_k^2 is below some threshold θ . Otherwise, image is unknown to the system and can be added to the database.

As it seen from the Fig. 5 image is treated as a vector. The coordinates can be regarded as a compact image representation.

This representation is not lossless and the exact value of loss depends on the basis. So, the more precise the basis will be the less noise the representations will contain and different images will more likely to have different coordinates. Our hypothesis is that the basis having smaller average loss over images will show greater recognition performance. The derivation of this loss function is given below.

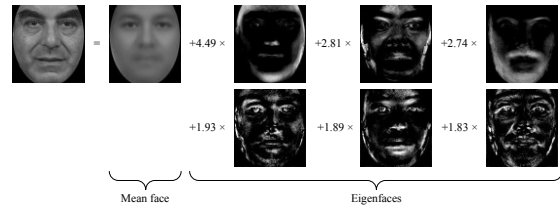


Fig. 5 – Image decomposition to mean face and eigenfaces with biggest values

Let A be a set of faces which are used to construct basis. Let B be a set of faces which are used to test basis. $|A|, |B|$ are the numbers of images in defined sets respectively. Let $L(v)_i$ be a loss vector:

$$L(v)_i = (B_i - B'(v)_i)^2, \quad (6)$$

where $B'(v)_i$ – face images restored from basis, v – basis:

$$B'(v)_i = (v^T W_i) + \Psi, \quad (7)$$

where W_i – representation of i th image in the basis v , Ψ – mean face.

The target function can be written down as follows:

$$e(v) = \frac{1}{|B|} \sum_{k=1}^{|B|} L(v)_k. \quad (8)$$

The task to minimize $e(v)$ is practically impossible. However, this target function can be used to compare existing eigenface bases. To

illustrate this idea let us build two different bases: v_{A1} consists of people of all races, gender, with or without mustaches (image set A1); v_{A2} – white male faces without mustache (image set A2). Each set of images consists of 100 images.

As expected, mean face computed from the first set of images (see Fig. 6, A) contains features of all races that were included into initial set of images. Mean face computed from the second image set is a face of a white male (see Fig. 6, B). As a test set we used a set of 45 images that consists of males and females of different races. Images from this set are not present in set A1 and A2.

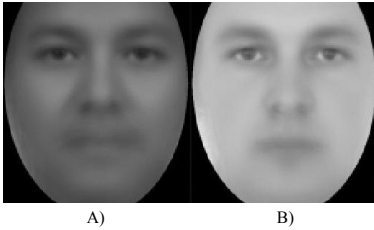


Fig. 6 – Two average faces

The following results were obtained: $e(v_{A1}) = 232587$ and $e(v_{A2}) = 421466$. Since $e(v_{A1}) < e(v_{A2})$ we can expect basis v_{A1} to restore faces more precisely than basis v_{A2} . Other experimental results are given in section 0.

6. GENDER DETERMINATION

One of the goals of our system is to handle indeed large MRTD databases. Although data storage optimizations, eager evaluation and repeated calculations caching reduce search time, full scan of the database takes much time. In order to obtain the preliminary result earlier, we divide the database into males and females using AdaBoost classifier [7]. It determines the gender of a particular person using only relatively small number of pixel comparisons [9]. Detailed description of this approach can be found in [12]. We will outline the main steps of the algorithm.

We extract faces so that the distance between the eyes is 10 pixels, pupils of the eyes are situated in points (5,5) and (10,5). The whole face must occupy the area with dimensions of 20×20 pixels. We use five types of comparisons along with their inverses. All together they represent weak classifiers:

1. $p_i > p_j$
2. $|p_i - p_j| \leq 5$
3. $|p_i - p_j| \leq 10$
4. $|p_i - p_j| \leq 25$
5. $|p_i - p_j| \leq 50$

where p_i, p_j are two pixels in the image, $i, j = \overline{1, 400}, i \neq j$.

Although the total quantity of weak classifiers in the image with dimensions 20×20 is 1596000, AdaBoost is capable of selecting and combining a small number of weak classifiers into the strong classifier, having significant accuracy. The first features selected by AdaBoost are shown in Fig. 7.

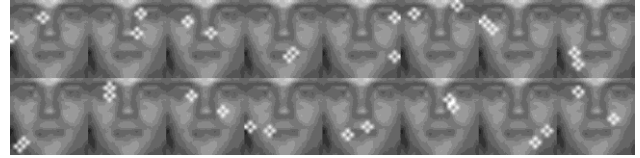


Fig. 7 – First pixel pairs (weak classifiers) selected by AdaBoost

We have configured AdaBoost to select 1000 features and combine them into one strong classifier. The dependence of classification accuracy on the number of weak classifiers is given in the Fig. 8.

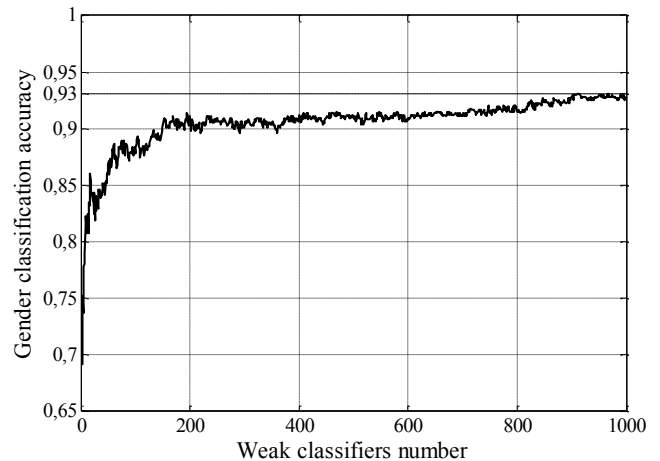


Fig. 8 – Accuracy of gender classifiers consisting from 1 to 1000 weak classifier. The highest accuracy is achieved with 911 weak nodes.

The best classifier consists of 911 nodes and gives 0,93 accuracy. We use this classifier to split the database into two groups and to determine gender of the unknown individual. Even though these percentiles are rather high, we cannot guarantee 100% accuracy. Because of that we use grouping of the individuals in the database to receive the preliminary searching result as fast as possible. This preliminary result in 93% cases will be identical to the final result. For example, if an individual is male, the system will start search in males group outputting results in real time. Due to small error equal to 0.07 it will proceed to search within females group, but the interim results will obtain with spending up to 40% less time.

7. EXPERIMENTAL RESULTS

We used frontal faces in the COLOR FERET[15][16] database to build bases. We used the rest of the FERET database (those images that were

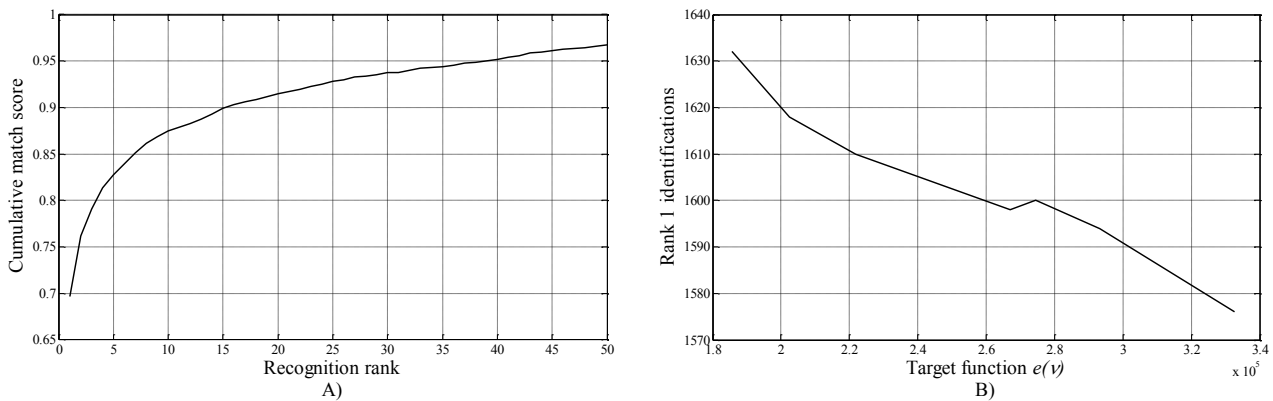


Fig. 9 – Recognition system performance (A), computed with the use of 11886 images. The dependence of the number of rank 1 true positives on the value of target function (B), computed with the use of 1983 images from the FERET database

not used to create bases) with additional 9903 frontal facial images to test the system. The total number of images made up 11886. We tried to locate individuals with facial expressions different to query image. In our previous work [17] we achieved 0.3 seconds per query over 11886 images database, which is quite promising. After splitting the database into two groups the preliminary results are ready after 0.18 seconds.

The cumulative match score of the system is shown on the Fig. 9 (A). The recognition rank at the position N denotes the ratio of correctly identified individuals that were placed on the N -th position by similarity to the query image.

Fig. 9 (B) shows the dependence of the number of rank 1 true positives upon the $e(v)$, derived in the previous section. Since relative values do not show significant difference, the graph shows absolute values. However, when the recognition system is to be applied in the critical environment, e.g. criminals recognition, it is worth tuning the basis to produce as much rank 1 identifications as possible. Computation of the exact number of rank 1 identifications is time consuming, while computing $e(v)$ is very fast.

8. CONCLUSIONS AND FUTURE RESEARCHES

The described system was implemented on the C++ programming language with the use of Open Source Computer Vision Library.

Variable image scale parameter helps increase speed and detection accuracy, allowing the system handling large images databases. Image processing sequence brings each face to the same state regardless of initial rotation and size. The proposed “the biggest face” approach along with regions that most likely contain eyes significantly increase the speed of feature detection.

The proposed method of basis ranking allows choosing the best one which is useful for more successful recognition process.

Gender-based identification allows getting preliminary results taking up to 40% less time.

One of the future research directions is to use additional criteria to create more groups in the database based on race, age, or face geometry.

9. ACKNOWLEDGMENT

Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office.

10. REFERENCES

- [1] M. Yang, Recent advances in face detection, *Tutorial, IEEE International Conference on Pattern Recognition*, Cambridge, 2004.
- [2] E. Osuna, R. Freund and F. Girosi, *Support Vector Machines: Training and Applications*, Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences, 1997.
- [3] S. Romdhani, P. Torr, B. Scholkopf and A. Blake, Computationally efficient face detection, *In Proceeding of the 8th International Conference on Computer Vision*, vol. 2, Vancouver, 2001, pp. 695-700.
- [4] H. Schneiderman, *A Statistical Approach to 3D Object Detection Applied to Faces and Cars*, Dissertation Thesis at Carnegie Mellon University, Pittsburgh, 2000.
- [5] H. Rowley, S. Baluja and T. Kanade, Neural network-based face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (20) 1 (1998), pp. 22-38.
- [6] P. Viola and M. Jones, Robust real-time face

- detection, *International Journal of Computer Vision*, (2004), pp. 137-154.
- [7] Y. Freund and R. Shapire, A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, (1996) pp. 119-139.
- [8] M. Turk, and A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience*, (3) 2 (1991), pp. 71-86.
- [9] S. Baluja, M. Sahami, and H. Rowley, Efficient face orientation discrimination, *International Conference on Image Processing*, 2004.
- [10] B. Moghaddam and M. Yang, Learning gender with support faces, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, (24) 5 (May 2002).
- [11] G. Shakhnarovich, P. Viola and B. Moghaddam, A unified learning framework for real time face detection and classification, *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002
- [12] S. Baluja and H. Rowley, Boosting sex identification performance, *International Journal of Computer Vision*, (71) 1 (2007).
- [13] *International Civil Aviation Organization. Machine Readable Travel Documents*, Doc 9303, 2006.
- [14] R. Lienhart and J. Maydt, An extended set of Haar-like features for rapid object detection, *In Proceedings of The IEEE International Conference on Image Processing*, pp. 900-903.
- [15] P.J. Phillips, H. Moon, S.A. Rizvi and P.J. Rauss, The FERET evaluation methodology for face recognition algorithms, *IEEE Trans. Pattern Analysis and Machine Intelligence*, (22) (2000), pp. 1090-1104.
- [16] P.J. Phillips, H. Wechsler, J. Huang and P. Rauss, The FERET database and evaluation procedure for face recognition algorithms, *Image and Vision Computing J.*, (16) 5 (1998), pp. 295-306.
- [17] S. Tulyakov and R. Sadykhov, Face recognition on machine readable travel documents: algorithms and results, *In Proceedings of International Conference on Pattern Recognition and Information Processing*, 2011.



Sergey Tulyakov, received his BS degree with honors in computer science from Belarusian State University of Informatics and Radioelectronics in 2009. In 2010 received master degree from the aforementioned university. He is currently a PhD student at Belarusian State University of Informatics and Radioelectronics. His research areas include image processing, computer vision and machine learning.



Rauf Kh. Sadykhov, habilitated doctor in computer science since 1991. Vice-chairman of the Belarusian office of the International Society of neural networks, IEEE member, member of the Belarusian Association of Image Processing. His research directions include: digital signal processing, pattern recognition and image processing, artificial intelligence, machine learning, parallel architectures for digital signal and image processing. He is an author and co-author of more than 360 scientific papers, including 3 monographs. Currently he is a head of Computer Systems Department at Belarusian State University of Informatics and Radioelectronics.