Gender Classification for Real-Time Audience Analysis System

Vladimir Khryashchev, Lev Shmaglit, Andrey Shemyakov, Anton Lebedev
Yaroslavl State University
Yaroslavl, Russia
vhr@yandex.ru, shmaglit_lev@yahoo.com, andrey.shemiakov@gmail.com,
lebedevdes@gmail.com

Abstract—**The system allowing to extract all the possible information about depicted people from the input video stream is discussed. As reported previously, the proposed system consists of five consecutive stages: face detection, face tracking, gender recognition, age classification and statistics analysis. The crucial part of the system is gender classifier construction on the basis of machine learning methods. We propose a novel algorithm consisting of two stages: adaptive feature extraction and support vector machine classification. Both training technique of the proposed algorithm and experimental results acquired on a large image dataset are presented. More than 90% accuracy of viewer's gender recognition is achieved.**

I. Introduction

Automatic video data analysis is a very challenging problem. In order to find a particular object in a video stream and automatically decide if it belongs to a particular class one should utilize a number of different machine learning techniques and algorithms, solving object detection, tracking and recognition tasks [1-3]. A lot of different algorithms, using such popular techniques as principal component analysis, histogram analysis, artificial neural networks, Bayesian classification, adaptive boosting learning, different statistical methods, and many others, have been proposed in the field of computer vision and object recognition over recent years. Some of these techniques are invariant to the type of analyzed object, others, on the contrary, are utilizing aprioristic knowledge about a particular object type such as its shape, typical color distribution, relative positioning of parts, etc. [4]. In spite of the fact that in the real world there is a huge number of various objects, a considerable interest is being shown in the development of algorithms of analysis of a particular object type – human faces. The promising practical applications of face recognition algorithms can be automatic number of visitors calculation systems, throughput control on the entrance of office buildings, airports and subway; automatic systems of accident prevention, intelligent human-computer interfaces, etc.

Gender recognition, for example, can be used to collect and estimate demographic indicators [5-8]. Besides, it can be an important preprocessing step when solving the problem of person identification, as gender recognition allows twice to reduce the number of candidates for analysis (in case of identical number of men and women in a database), and thus twice to accelerate the identification process.

In order to organize a completely automatic system, classification algorithms are utilized in the combination with a face detection algorithm, which selects candidates for further analysis [9-14]. In paper [15] we proposed a system which extracts all the possible information about depicted people from the input video stream, aggregates and analyses it in order to measure different statistical parameters (Fig. 1).

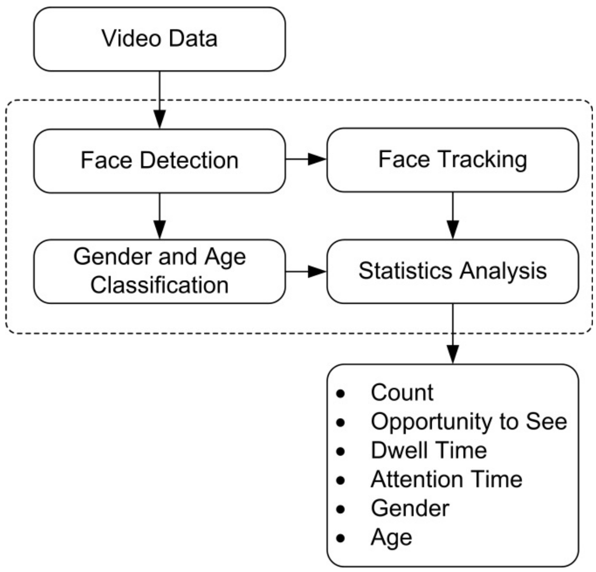

Fig. 1. A block diagram of the proposed application for video analysis

The quality of face detection step is critical to the final result of the whole system, as inaccuracies at face position determination can lead to wrong decisions at the stage of recognition. To solve the task of face detection AdaBoost classifier, described in paper [16], is utilized. Detected fragments are preprocessed to align their luminance characteristics and to transform them to uniform scale.

On the next stage detected and preprocessed image fragments are passed to the input of gender recognition classifier which makes a decision on their belonging to one of two classes («Male», «Female»). Same fragments are also analyzed by the age estimation algorithm which divides them into several age groups [17]. Age classifier shares the same learning technique with that of gender classifier.

As a result the following metrics are calculated:

- Count — the number of viewers who've paid an attention to a particular product or have watched the advertisement.
- Opportunity to See (OTS) — the number of potential viewers who were close to the presented product or advertising media;
- Dwell Time — the average time during which potential viewers have been in the visibility range to the presented product or advertising media;
- Attention Time — the average time when the viewer was watching the object of interest;
- Gender — viewer gender (man/woman);
- Age — viewer age group (child / youth / adult / seniors).

This paper focuses on the problem of gender recognition classifier construction. There are comprehensive surveys written on face detection [14, 8], face recognition [18] and facial expression analysis [19]. Gender recognition has been studied less. The comparative analysis of lately proposed gender classification algorithms has been presented in paper [3]. The researches have proposed algorithms based on artificial neural networks [4], on a combination of Gabor wavelets and principal component analysis [20], on independent component analysis and linear discriminant analysis (LDA) [21, 22]. In paper [23] the authors utilize genetic algorithm for feature selection and support vector machine (SVM) [24] for classification. In paper [25] an algorithm based on local binary patterns in combination with AdaBoost classifier was proposed. Experiments with radial basis function (RBF) networks and inductive decision trees are described in paper [26].

A novel gender recognition algorithm, proposed in this paper, is based on non-linear SVM classifier with RBF kernel. To extract information from image fragment and to move to a lower dimension feature space we propose an adaptive feature generation algorithm which is trained by means of optimization procedure according to LDA principle.

The rest of the paper is organized as follows. The scheme of the proposed algorithm is described in section 2. Section 3 considers training methodology and experimental setup. In section 4, the results of the proposed algorithm

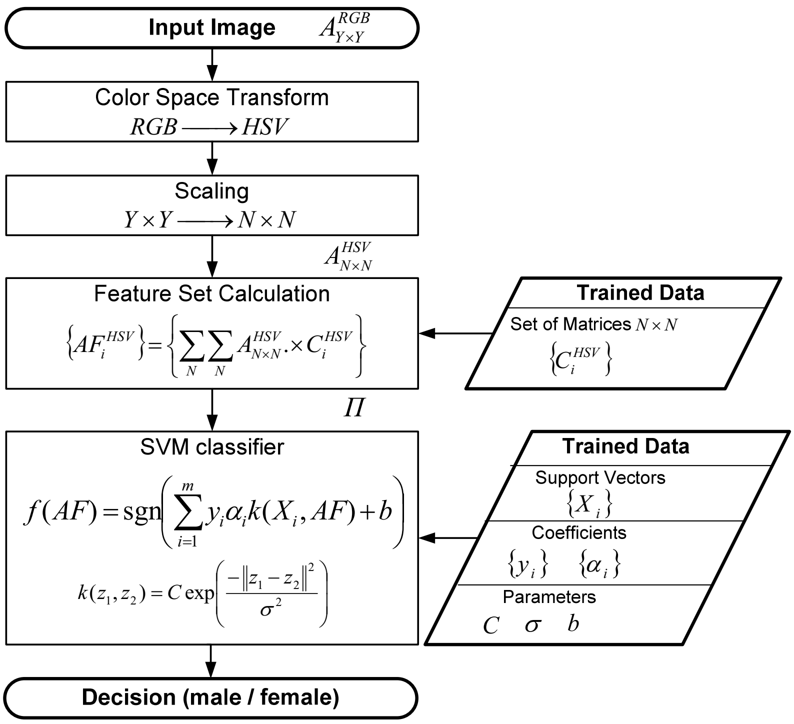

Fig. 2. The scheme of the proposed gender classification algorithm

comparison with state-of-the-art gender classification methods is presented. Section 5 concludes the paper.

II. The Proposed Algorithm

Classifier is based on Adaptive Features and SVM (AF-SVM). Its operation includes several stages, as shown in Fig. 2.

AF-SVM algorithm consists of the following steps: color space transform, image scaling, adaptive feature set calculation and SVM classification with preliminary kernel transformation. Input image $A_{Y\times Y}^{RGB}$ is converted from RGB to HSV color space and is scaled to fixed image resolution $N \times N$. After that we calculate a set of features $\{AF_i^{HSV}\}$, where each feature represents the sum of all rows and columns of element-by-element matrix product of an input image and a coefficient matrix C_i^{HSV} with resolution $N \times N$, which is generated by the training procedure:

$$AF_i^{HSV} = \sum_N \sum_N A_{N\times N}^{HSV} . \times C_i^{HSV} .$$

The obtained feature vector is transformed using a Gaussian radial basis function kernel:

$$k(z_1, z_2) = C \exp\left(\frac{-\|z_1 - z_2\|^2}{\sigma^2} \right).$$

Kernel function parameters C and σ are defined during training. The resulted feature vector serves as an input to linear SVM classifier which decision rule is:

$$f(AF) = \operatorname{sgn}\left(\sum_{i=1}^{m} y_i \alpha_i k(X_i, AF) + b \right).$$

The set of support vectors $\{X_i\}$, the sets of coefficients $\{y_i\}$, $\{\alpha_i\}$ and the bias b are obtained at the stage of classifier training.

III. Training and Testing Setup

Both gender recognition algorithm training and testing require huge enough color image database. The most commonly used image database for the tasks of human faces recognition is the FERET database [27], but it contains insufficient number of faces of different individuals, that's why we collected our own image database, gathered from different sources (Table 1). Faces on the images from the proposed database were detected automatically by AdaBoost face detection algorithm. After that false detections were manually removed, and the resulted dataset consisting 10 500 image fragments (5 250 for each class) was obtained. This dataset was split into three independent image sets: training, validation and testing. Training set was utilized for feature generation and SVM classifier construction. Validation set was required in order to avoid the effect of overtraining during the selection of optimal parameters for the kernel function. Performance evaluation of the trained classifier was carried out with the use of the testing set.

TABLE I. THE PROPOSED TRAINING AND TESTING IMAGE DATABASE PARAMETERS

Parameter	Value
The total number of images	8 654
The number of male faces	5 250
The number of female faces	5 250
Minimum image resolution	640×480
Color space format	RGB
Face position	Frontal
People's age	From 18 to 65 years old
Race	Caucasian
Lighting conditions, background and facial expression	No restrictions

The training procedure of the proposed AF-SVM classifier can be split into two independent parts: feature generation, SVM construction and optimization. Let's consider the feature generation procedure. It consists of the following basic steps:

- RGB → HSV color space transform of the training images (all further operations are carried out for each color component independently);
- scaling training images to fixed image resolution $N \times N$;
- coefficient matrix C_i^{HSV} random generation;
- feature value AF_i^{HSV} calculation for each training fragment;
- the utility function F calculation as a square of a difference between feature averages, calculated for "male" and "female" training image datasets, divided by the sum of feature variances [28]:
- $F = \dfrac{\left(\left\langle \{AF_i^{HSV}\}_M \right\rangle - \left\langle \{AF_i^{HSV}\}_F \right\rangle \right)^2}{\sigma\{AF_i^{HSV}\}_M + \sigma\{AF_i^{HSV}\}_F}$;
- iteratively in a cycle (until the number of iterations exceeds some preliminary fixed maximum value): random generation of coefficient matrix $\widetilde{C}_i^{HSV}$ inside the fixed neighborhood of matrix C_i^{HSV}, feature value $A\widetilde{F}_i^{HSV}$ calculation for each training fragment, calculation of the utility function $\widetilde{F}$, transition to a new point ($F \to \widetilde{F}$, $C \to \widetilde{C}$), if $\widetilde{F} > F$;

- saving the matrix C_i^{HSV} after exceeding the maximum number of iterations;
- return to beginning in order to start the generation of the next (i+1) feature.

An optimization procedure, described above, allows to extract from an image only the information which is necessary for class separation. Besides, features with higher utility function value have higher separation ability. The feature generation procedure have the following adjusted parameters: training fragments resolution (*N*), the number of training images for each class (*M*), maximum number of iterations (*T*). The following values, as a compromise between reached separation ability and the training speed, were empirically received:

$$N = 65; \qquad M = 400; \qquad T = 10^5.$$

1000 features have been generated for each color component. At the second stage of training these features have been extracted from training images and were then used to learn an SVM classifier. SVM construction and optimization procedure included the following steps:

- calculation of the feature set, generated on the first stage of training, for each training fragment;
- feature normalization;
- learning an SVM classifier with different parameters of the kernel function;
- recognition rate (RR) calculation using validation image dataset;
- determination of optimal kernel function parameters (maximizing RR);
- learning a final SVM classifier with the found optimal kernel function parameters.

The goal of SVM optimization procedure is to find a solution with the best generalization ability, and thus with the minimum classification error. The adjusted parameters are: the number of features in a feature vector (*N2*), the number of training images for each class (*M2*), the kernel function parameters σ and C.

Grid search was applied to determine optimal kernel parameters: SVM classifier was constructed varying $C = 10^{k1}$ and $\sigma = 10^{k2}$, where $k1$ and $k2$ – all combinations of integers from the range [-15 … 15]; during the search recognition rate was measured using validation image dataset. The results of this procedure are presented in Fig. 3. Maximum recognition rate (about 80%) was obtained for $C = 10^6$ and $\sigma = 10^8$.

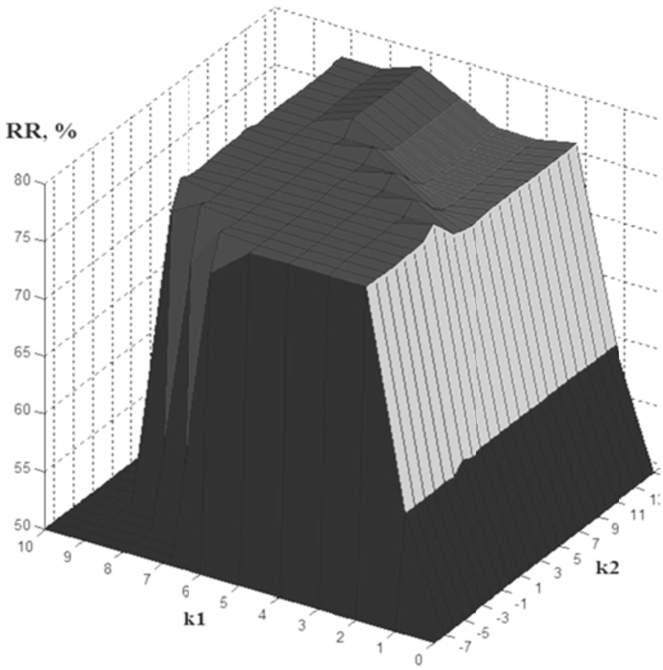

Fig. 3. The dependence of RR from kernel function parameters $C = 10^{k1}$ and $\sigma = 10^{k2}$

Besides, we investigated the dependence of classifier

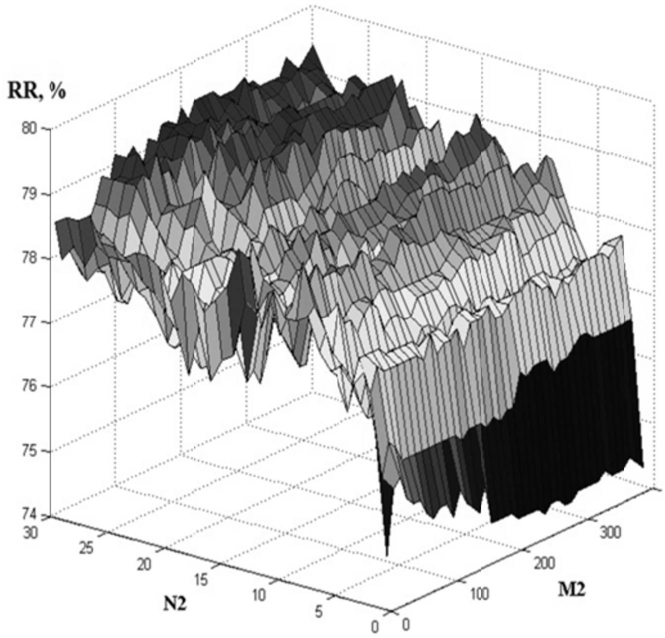

Fig. 4. The dependence of recognition rate from training procedure parameters

performance from the number of features extracted from each color component – *N2*, and from the number of training images for each class – *M2* (Fig. 4).

The analysis shows that each feature has essential separation ability, and at $N2 = 30$ recognition rate reaches 79.5%. At the same time the growth of RR is observed both with the growth of $N2$ and $M2$ due to the accumulation of information about considered classes inside the classifier. Thus, to obtain a compromise between quality and speed the following parameters were chosen: $N2 = 30$; $M2 = 400$.

IV. Experimental Results

In this section we present the results of the proposed AF-SVM algorithm comparison with state-of-the-art classifiers: SVM [24] and KDDA (Kernel Direct Discriminant Analysis) [21].

Classifier AF-SVM was trained according to a technique, given above. SVM and KDDA classifiers have far less adjustable parameters as they are working directly with image pixel values instead of feature vectors. To construct these classifiers the same training base, as for AF-SVM classifier, was used. The following conditions also were identical for all three considered classifiers: the number of training images for each class, training fragments resolution and image preprocessing procedure. Optimization of SVM and KDDA kernel function parameters was held using the same technique and the same validation image dataset as used in case of AF-SVM classifier. Thus, equal conditions for independent comparison of considered classification algorithms, using testing image dataset, were provided.

For the representation of classification results we utilized the Receiver Operator Characteristic (ROC-curve) [29]. As there are two classes, one of them is considered to be a positive decision and the other – a negative. ROC-curve is created by plotting the fraction of true positives out of the positives (TPR = true positive rate) vs. the fraction of false positives out of the negatives (FPR = false positive rate), at various discrimination threshold settings. The advantage of ROC-curve representation lies in its invariance to the relation between the first and the second error type's costs.

TABLE II. Comparative Analysis of Tested Algorithms Performance

Algorithm / **Parameter**	**SVM**		**KDDA**		**AF-SVM**	
Recognition rate	True	False	True	False	True	False
Classified as "male", %	80	20	75.8	24.2	80	20
Classified as "female", %	75.5	24.5	65.5	34.5	79.3	20.7
Total classification rate, %	**77.7**	22.3	**69.7**	30.3	**79.6**	20.4
Operation speed, faces/sec	44		45		65	

The results of AF-SVM, SVM and KDDA testing are presented in Fig. 5 and in table II. The computations were held on a personal computer with the following configuration: operating system – Microsoft Windows 7; CPU type – Intel Core i7 (2 GHz) 4 cores; memory size – 6 GB.

The analysis of testing results show that AF-SVM is the

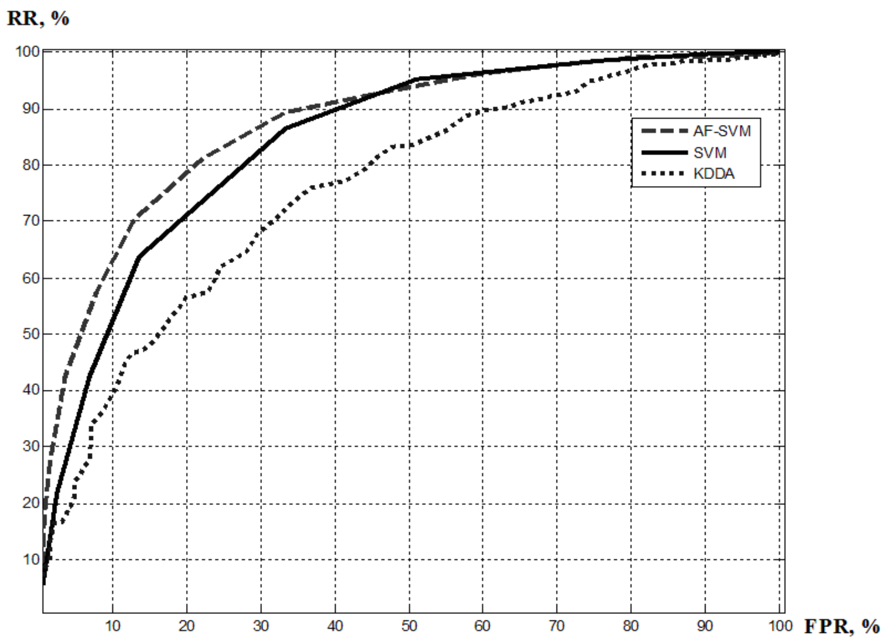

Fig. 5. ROC-curves of tested gender recognition algorithms

most effective algorithm considering both recognition rate and operational complexity. AF-SVM has the highest RR among all tested classifiers – 79.6% and is faster than SVM and KDDA approximately by 50%.

Such advantage is explained by the fact that AF-SVM algorithm utilizes a small number of adaptive features, each of which carries a lot of information and is capable to separate classes, while SVM and KDDA classifiers work directly with a huge matrix of image pixel values.

Let's consider the possibility of classifier performance improvement by the increase of the total number of training images per class from 400 to 5000. Experiments showed that SVM and KDDA recognition rates can't be significantly improved in that case. Besides, their computational complexity increases dramatically with the growth of the training dataset. This is explained by the fact that while the number of pixels, which SVM and KDDA classifiers utilize to find an optimal solution in a high dimensional space, increases it becomes harder and even impossible to find an acceptable solution for the reasonable time.

In the case of AF-SVM classifier the problem of the decrease of SVM classifier efficiency with the growth of the training database can be solved by the use of the small number of adaptive features, holding information about a lot of training images at once. For this purpose we suggest that each feature should be trained using a random subset

TABLE III. RECOGNITION RATE OF AF-SVM ALGORITHM TRAINED ON DATASETS OF DIFFERENT SIZE

Algorithm / Parameter	AF-SVM (M=5000)		AF-SVM (M=400)	
Recognition rate	True	False	True	False
Classified as "male", %	90.6	9.4	80	20
Classified as "female", %	91	9	79.3	20.7
Total classification rate, %	**90.8**	9.2	**79.6**	20.4

(containing 400 training images per class) from the whole training database (containing 5000 images per class). Thus each generated feature will hold the maximum possible amount of information, required to divide the classes, and a set of features will include the information from each of the 10 000 training images.

On the stage of feature generation 300 features were trained according to the technique described above. After that an SVM classifier, utilizing these features, was trained similarly as before. Besides, we preserved the number of training images for SVM construction equal to 400, and thus the operation speed of the final classifier remained the same as in previous experiments – 65 faces processed per second.

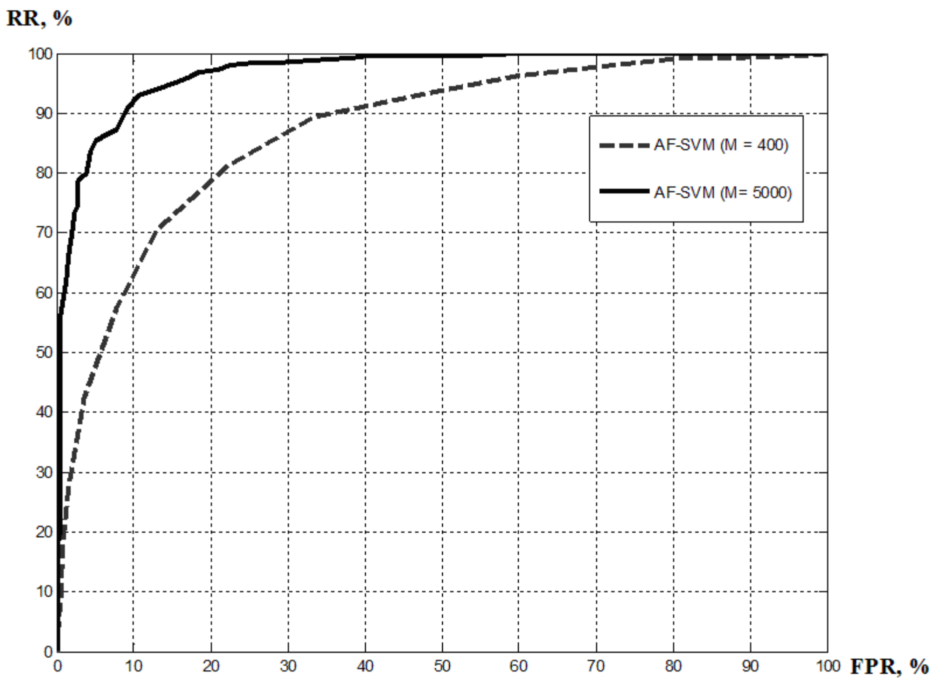

Fig. 6. ROC-curves for AF-SVM algorithm trained on datasets of different size

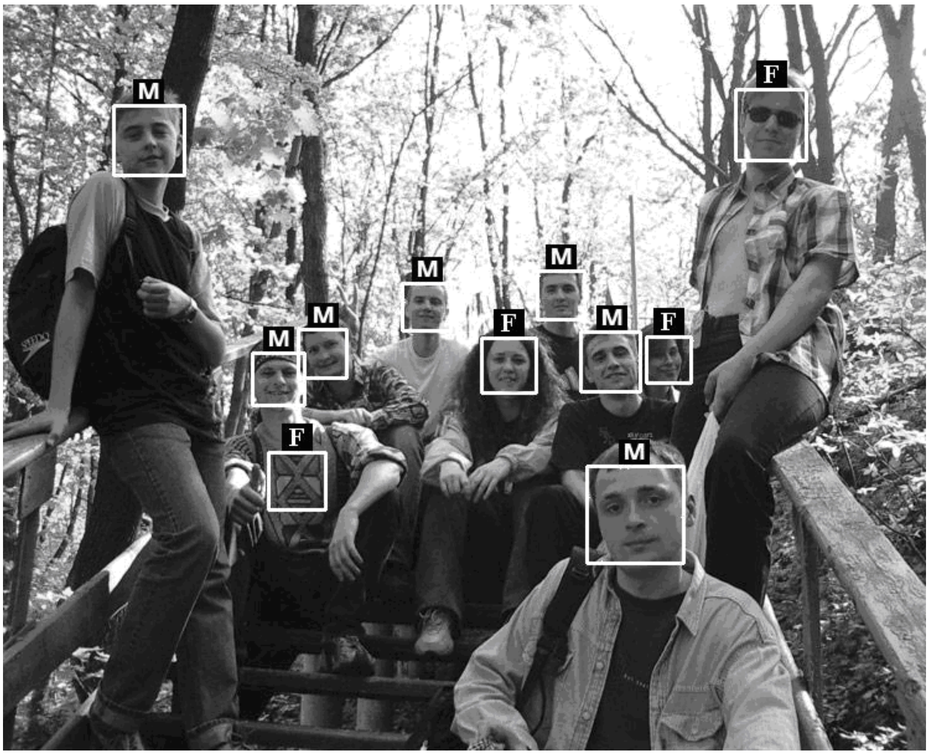

Fig. 7. Visual example of the proposed automatic gender classifier

The results of AF-SVM algorithm trained using expand dataset ($M = 5000$) and the initial AF-SVM classifier ($M = 400$) comparison are presented in table III and in Fig. 6.

The results show that AF-SVM algorithm together with the proposed training setup allow to significantly improve the classifier performance in case of increasing the training database size to 5000 images per class. Recognition rate of nearly 91% is achieved. Visual example of the proposed gender classification system is presented in Fig. 7. Recognized classes “male” and “female” are designated by symbols “M” and “F” correspondingly (Fig. 7).

V. CONCLUSION

A new classifier based on adaptive features and support vector machines, solving the problem of automatic gender recognition via face area analysis, is proposed. It shows recognition rate of 79.6%, which is 1.9% and 9.9% higher than that of SVM and KDDA correspondingly. The possibility of AF-SVM recognition rate improvement up to 91% in case of increasing the training database size to 5000 images per class is shown. The proposed algorithm allows to process 65 faces per second which is enough to use it in real time video sequence analysis applications.

The adaptive nature of the feature generation procedure allows using the proposed AF-SVM classifier for the recognition of any other object on an image (in addition to faces). For this purpose a new training database of the considered class should be constructed and a classifier should be retrained according to the described in this paper technique.

The proposed gender recognition classifier is used as a part of video data analysis system, which provides collection and processing of information about the audience in real time. The system is fully automatic and does not require people to conduct it. No personal information is saved during the process of operation. The noted features allow applying the proposed system in various spheres of life: places of mass stay of people (stadiums, movie theaters and shopping centers), transport knots (airports, railway and auto stations), border passport and visa control check-points, etc.

It should be noted that this report covers mostly theoretical aspects of gender classifier construction while the performance of the proposed audience analysis system will be demonstrated at FRUCT conference demo section.

This work was supported by P.G. Demidov Yaroslavl State University research framework (project #1060).

REFERENCES

[1] E. Alpaydin, *Introduction to Machine Learning*. The MIT Press, 2010.

[2] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer, 2010.

[3] E. Makinen, R. Raisamo, "An experimental comparison of gender classification methods," *Pattern Recognition Letters*, vol. 29, No 10, pp. 1544-1556, 2008.

[4] S. Tamura, H. Kawai, H. Mitsumoto, "Male/female identification from 8 to 6 very low resolution face images by neural network," *Pattern Recognition Letters*, vol. 29, No 2, pp. 331-335, 1996.

[5] V. Khryashchev, A. Priorov A, L. Shmaglit, M. Golubev, "Gender Recognition via Face Area Analysis," *Proc. World Congress on Engineering and Computer Science, Berkeley, USA. P. 645-649*, 2012.

[6] V. Khryashchev, A. Ganin, M. Golubev, L. Shmaglit, "Audience analysis system on the basis of face detection, tracking and classification techniques," *Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2013*, 13-15 March, 2013, Hong Kong, pp. 446-450.

[7] Y. Fu, T.S. Huang, "Age Synthesis and Estimation via Faces: A Survey," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, No 11, pp. 1955-1976, 2010.

[8] E.C. Sung, J.L. Youn, J.L. Sung, R.P. Kang and K. Jaihie, "A comparative study of local feature extraction for age estimation," *IEEE Int. Conf. on Control Automation Robotics & Vision (ICARCV)*, pp. 1280-1284, 2010.

[9] G. Guodong, M. Guowang, "Human age estimation: What is the influence across race and gender?" *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 71-78, 2010.

[10] L. Zhen, F. Yun, T.S. Huang, "A robust framework for multiview age estimation," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 9-16, 2010.

[11] G. Guodong, W. Xiaolong, "A study on human age estimation under facial expression changes," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 2547-2553, 2012.

[12] L.W. Hee, W. Jian-Gang, Y. Wei-Yun, L.C. Xing, P.T. Yap, "Effects of facial alignment for age estimation," *IEEE Int. Conf. on Control Automation Robotics & Vision (ICARCV)*, pp. 644-647, 2010.

[13] D. Kriegman, M.H. Yang, N Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, No 1, pp. 34-58, 2002.

[14] E Hjelmas, "Face detection: A Survey," *Computer vision and image understanding*, vol. 83, No 3, pp. 236-274, 2001.

[15] V. Pavlov, V. Khryashchev, E. Pavlov, L. Shmaglit, "Application for Video Analysis Based on Machine Learning and Computer Vision Algorithms," *Proceedings of the 14th conference of open innovation association FRUCT*, pp. 90–100, 2013.

[16] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 511-518, 2001.

[17] P. Thukral, K. Mitra, R. Chellappa, "A hierarchical approach for human age estimation," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1529-1532, 2012.

[18] W. Zhao, R. Chellappa, P. Phillips, A Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys (CSUR)*, vol. 35, No 4, pp. 399-458, 2003.

[19] B. Fasel, J. Luettin, "Automatic Facial Expression Analysis: A Survey," *Pattern Recognition Letters*, vol. 36, No. 1, pp. 259–275, 2003.

[20] M. Lyons, J. Budynek, A. Plante, S. Akamatsu, "Classifying Facial Attributes Using a 2-d Gabor Wavelet Representation and Discriminant Analysis," *Proc. Int. Conf. on Automatic Face and Gesture Recognition*, pp. 202–207, 2000.

[21] A. Jain, J. Huang, "Integrating Independent Components and Linear Discriminant Analysis for Gender Classification," *Proc. Int. Conf. on Automatic Face and Gesture Recognition*, pp. 159–163, 2004.

[22] Y. Saatci, C. Town, "Cascaded Classification of Gender and Facial Expression Using Active Appearance Models," *Proc. Int. Conf. on Automatic Face and Gesture Recognition*, pp. 393–400, 2006.

[23] Z. Sun, G. Bebis, X. Yuan, S.J. Louis, "Genetic Feature Subset Selection for Gender Classification: A Comparison Study," *Proc. IEEE Workshop on Applications of Computer Vision*, pp. 165–170, 2002.

[24] C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.

[25] N. Sun et al, "Gender Classification Based on Boosting Local Binary Pattern," *Proc. Int. Symposium on Neural Networks*, vol. 2, pp. 194–201, 2006.

[26] S. Gutta, H. Wechsler, P.J. Phillips, "Gender and Ethnic Classification of Face Images," *Proc. Int. Conf. on Automatic Face and Gesture Recognition*, pp. 194–199, 1998.

[27] P.J. Phillips et al, "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[28] H. Gao, J. Davis, "Why Direct LDA is not Equivalent to LDA," *Pattern Recognition Letters*, vol. 39, no. 5, pp. 1002–1006, 2006.

[29] T. Fawcett, "ROC Graphs: Notes and Practical Considerations for Researchers," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 882–891, 2004.