Improving the Face Gender Classification by the Set of Features

Vladimir Khryashchev, Vladimir Pavlov, Andrey Priorov
P.G. Demidov Yaroslavl State University
Yaroslavl, Russia
vhr@yandex.ru, vladimir@1pavlov.com, andcat@yandex.ru

Igor Kosterin
Fire and Rescue Academy
Ivanovo, Russia
kosteriniv@gmail.com

***Abstract*—Gender recognition using face images is one of interesting for practical applications face analysis tasks. Most of the existing studies have focused on face images acquired under controlled conditions, such as famous FERET database. However, real applications require gender classification on real-life faces, which is much more challenging due to significant appearance variations in unconstrained scenarios. In this paper, we investigate gender recognition on real-life faces using the Labeled Faces in the Wild (LFW) dataset and our own RUS-FD dataset. We propose a gender classifier using three types of local features: Scale Invariant Feature Transform (SIFT) which is one of the most commonly-used ones because it is invariant to image scaling, translation and rotation, Histogram of Oriented Gradient (HOG) features, which is able to capture local shape information from the gradient structure with easily controllable degree of invariance to translations and the Gabor wavelets which reflect the multi-scale directional information. We obtain the performance of 96.79% by applying boosting learning on LFW dataset 92.04% by applying Support Vector Machine (SVM) on RUS-FD dataset. The approach proposed in this paper is promising to be further studied on other face classification tasks, such as age estimation and emotion recognition.**

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 [1-4]. 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. [14].

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 [9-15]. In paper [9] we proposed an audience measurement 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).

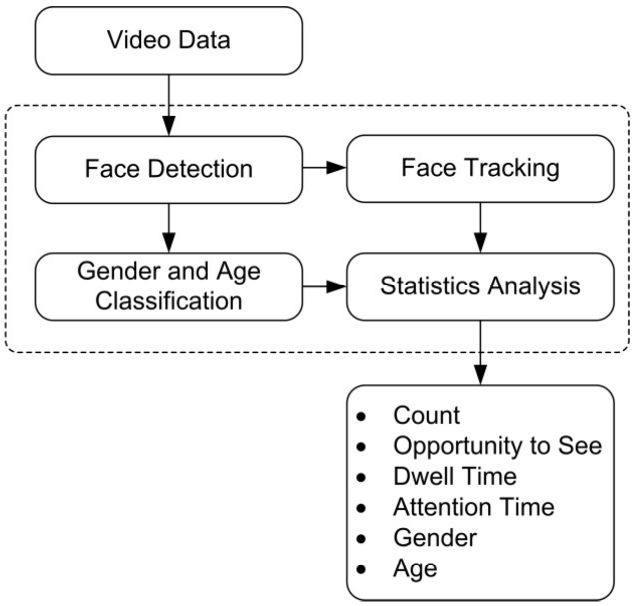

Fig. 1. A block scheme of the proposed application for video analysis

The following metrics are calculated by audience measurement system:

- Count — the number of viewers who've 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).

Gender recognition is a fundamental task for human beings, as many social functions critically depend on the correct gender perception [18-20]. Automatic gender classification has many important applications, for example, intelligent user interface, visual surveillance, collecting demographic statistics for marketing, etc. Human faces provide important visual information for gender perception. Gender classification from face images has received much research interest in the last two decades.

Different machine learning and mathematical statistics approaches are used for creating a gender classifier [24], [25], [27]. The facial image is a matrix of pixel values. There are many different approaches to determine how works with pixels and what to put to the input of a classifier. Sometimes to input put pixel values, sometimes emit features by using special algorithms. Some researchers use efficient features such as the Local Binary Patterns (LBP) [12] and Weber Local Descriptor (WLD) [17]. Others adopt more complicated features including the gradient information or wavelet functions. Among these features, the Scale Invariant Feature Transform (SIFT) is one of the most commonly-used ones because it is invariant to image scaling, translation, and rotation [21].

A new gender recognition algorithm, proposed in this paper, is based on non-linear Support vector machine (SVM) classifier [16] and has several types of features: SIFT, HOG, and Gabor filters. We use an AdaBoost algorithm for creating a classifier which works with features. To improve the algorithm, we use a Difference of Gaussian (DOG) [26] filter for pretreatment facial area.

The rest of the paper is organized as follows. Description of the principles of the features and the proposed algorithm described in the second part. An experimental comparison of the proposed algorithm with the existing competitors is presented in the third part for different training datasets: LFW and RUS-FD. Conclusions presented in the fourth part.

II. FACE GENDER CLASSIFICATION ALGORITHM

A new gender classification algorithm, proposed in this paper is based on non-linear SVM classifier and has several types of features calculations shortly described below.

A. The Scale-Invariant Feature Transform

The feature implemented in the vlfeat library [1]. Histogram computation size is 4 and the descriptor step is 16. For each block with step 16 to calculate the horizontal and vertical window descriptor 16x16 size.

B. Histogram of Oriented Gradients

Principle: the image is divided into 4 equal parts. In each part of the histogram calculated by 16 lines and formed bins in opposite directions one. Thus, it turns 8 destinations in 4 parts of the unit or the 32 factor. The result is 8 destinations in 4 parts of the unit or the 32 factor.

C. Gabor filters

This approach has been implemented to reduce the running time. The algorithm uses a filter with an aperture of 19x19. The convolution is performed via a Fourier transform. After receipt of the Gabor filter coefficients occurs decimation. The value is converted block size from 2x2 to 6x6, depending on the block size. When calculating the Gabor filter is performed linear decimation by 2 in both directions. Further decimation occurs from 1x1 nearest neighbor to a 3x3 size depending on the block.

D. Pre-selection of blocks

Blocks - is a part of the image where was selected the feature. Parts are selected randomly. In the next step are selected blocks which suitable for recognition by a AdaBoost algorithm [4], [7]. This approach is much faster than select in the first step part and after a feature.

For learning the algorithm 500 blocks was selected with each area 800 pixels. These blocks are used to calculate the characteristics of the type of HOG and GABOR. 300 blocks with each area 1200 pixels were selected for the SIFT features.

For each pair of features the feature vector f is calculated. Vector f form a matrix F for the entire training sample, where the line number corresponds the image number of the training sample. The vector y has 1 is the image class is "male" and -1 when the class is "female". The regression between the matrix F and vector y is calculated this information. To account for the displacement to the matrix F is added to the right column of the 1st class ($F_1 = [F, 1]$).

The vector of the regression coefficients is the solution of following equation:

$$(F_1' * F_1) * a = F_1' * y.$$

Pseudocode of proposed algorithm is represented in Fig. 2.

Vector x can be counted for each block and features, which is nearing y and equal $x = F_1 * a$. Each pair feature & block on the test set will match its vector, the number of test samples in sample length. The set of vector column is input of RealAdaBoost learning algorithm [1].

On the next step we need to select the partition of {-1,1} at intervals. Each element of the training sample x will correspond to a certain interval. Further, we know that the i-th test image belongs to the j-th interval, if $x(i)$ is included in this interval. In this embodiment the interval was selected in 32 parts. It was observed that as the number of intervals over 16 significantly increases the quality of the model, but slows the learning process.

T – the number of weak classifiers

Input: Training set

$\{(x_i; y_i)\}$, i = 1,...,N; $x_i \in R^d, y_i \in \{-1,1\}$.

1. Initialize false positive rate, classifier output and sample weight:

$$falsePos = 0.5, h = 0, weights = \frac{1}{N}$$

2. Repeat for t=1,2,…,T:

2.1 Update the sample weight:

$$weights = weights * e^{-y*h}, weights = \frac{weights}{\sum weights}$$

2.2 For m = 1 to M

2.2.1 Select cyclically SIFT, HOG, GABOR + index random block

2.2.2 Choose the correct vector x from X.

2.2.3 Calculate WP and WM for a vertor x.

2.2.4 For each m:

$$Z(m) = 2\sum_j \sqrt{WP_j WM_j} + CA * falsePos * C$$

2.3 Select the $feature + block$:

$$h(i) = 1/2 * \ln\left(\frac{WP_j + eps}{WM_j + eps}\right)$$

2.4 Update strong classifier:

$$F_{t+1}(x_i) = F_t(x_i) + h_t(x_i)$$

3. Output classifier

$$F(x) = sign\left(\sum_{j=1}^{T} h_j(x)\right)$$

Fig. 2. Pseudocode of proposed algorithm

To improve the classification we use different approaches for preprocessing facial area. In this experiment, we used an approach Difference of Gaussian (DOG) filter [26]. This filter allows selecting the most important information from the image. The principle of the algorithm involves the subtraction of one blurred version of an original image from another, less blurred version of the original. In the simple case of grayscale images, the blurred images are obtained by convolving the original grayscale images with Gaussian kernels having differing standard deviations. Blurring an image using a Gaussian kernel suppresses only high-frequency spatial information. The most important parameters for this algorithm is two version of the radius. They are the easiest to specify looking at viewer. It should be remembered that the increase in short-range leads to wide borders, and a decrease in long-range increases the limit on which is determined by the border or not. In most cases, the best results are obtained when the first radius smaller than the radius two. The resulting image is a blurred version of the original image.

As a result, the selected 25 features presented in pairs & block type vectors and 25 are input to the learning algorithm linear Support vector machine. For recognition of the input image on the floor necessary to extract 25 features of the image. Each feature scalar of each on its own vector of regression coefficients. On the 25 received features works especially SVM for gender recognition. Out model works with WP – the histogram for the elements of the vector x, which corresponding “male” in the interval (-1..1) with a step $\frac{2}{Nb}$. WM – the similar histogram only for “women”.

III. Experimental results

We conduct experiments on the two real-life dataset of images LFW [29] and RUS-FD [9]. LFW is a database for studying the problem of unconstrained face recognition, which contains 13,233 color face photographs of 5,749 subjects collected from the web. All the faces were detected by the Viola-Jones face detector [22], [23], [28], and the images were centered using detected faces and scaled to the size of 250x250 pixels. We manually labeled the ground truth regarding gender for each face. The faces that are not (near) frontal, as well as those for which it is dicult to establish the ground truth, were not considered. In our experiments, we chose 7,443 face images (2,943 females and 4,500 males); see Fig 3 for some examples. All experimental results were obtained using the 5-fold cross-validation. We partitioned the data set into five subsets of similar size, keeping the same ratio between female and male. The images of a particular subject appear only in one subset. The parameters of RUS-FD dataset is given at Table I. For the training set used by 80% of the images regardless of the type of image set.

TABLE I. The proposed RUS-FD training and testing image dataset parameters

Parameter	Value
The total number of images	6 000
The number of male faces	4 000
The number of female faces	2 000
Minimum image resolution	60×60
Color space format	RGB
Face position	Frontal
People’s age	From 18 to 65 years old
Race	Caucasian

The experimental results are given in Table II (with learning on LFW dataset) and Table III (with learning on RUS-FD dataset). In Table II there is also a comparison with approach based on local binary pattern (LBP) described in paper [30].

a) b)

Fig. 3. Visual examples of LFW-a dataset (a) and RUS-FD dataset (b)

TABLE II. EXPERIMENTAL RESULTS WITH LEARNING ON LFW DATASET

Approach						Recognition Rates (%)		
Features	Learning Database	Test Database	Pre-processing	Features Dimension	Classifier	Female	Male	Overall
Raw pixels	LFW	LFW	-	2944	SVM	86.89	94.13	91.27±1.67
Standart LBP	LFW	LFW	-	2478	SVM	89.78	95.73	93.38±1.50
Boosted LBP	LFW	LFW	-	500	Adaboost	91.98	95.98	94.40±0,86
Boosted LBP	LFW	LFW	-	500	SVM	92.02	96.64	94.81±1.10
HoG	LFW	LFW	DoG + contrast alignment		SVM	85.49	95.79	90.64
HoG	LFW	LFW	-		SVM	91.76	96.08	93.92
HoG+Gabor+SIFT	LFW	LFW	DoG + contrast alignment		Adaboost	89.51	97.12	93.32
HoG+Gabor+SIFT	LFW	LFW	-		Adaboost	95.65	97.92	96.79
HoG+Gabor+SIFT	LFW	RUS-FD	-		Adaboost	83.7	84.23	83.97

TABLE III. EXPERIMENTAL RESULTS WITH LEARNING ON RUS-FD DATASET

Approach					Recognition Rates (%)		
Features	Learning Database	Test Database	Pre-processing	Classifier	Female	Male	Overall
HoG	RUS-FD	LFW	DoG + contrast alignment	SVM	78.78	89.14	83.96
HoG	RUS-FD	RUS-FD	DoG + contrast alignment	SVM	91.48	92.60	92.04

Test results show that the both realization of the algorithm accuracy definition of "male" class (average result is better 2-3%). This result is caused by a large number of images in the training dataset for the first class. Also, this result confirms the theory about the high number of important a training data. The average accuracy of the floor at the maximum (when testing and training the same) is 96.79%. In a real-world conditions, where the training data set and test data dataset is different, the average accuracy is 83.96%. This result leads to the conclusion that the implementation of the algorithm can be used in real system.

IV. CONCLUSIONS

In this paper, we investigate gender classification on real-life faces acquired in unconstrained conditions, a challenging but relatively understudied problem. We learn an algorithms with a set of features (SIFT, Gabor and HOG) for gender classification. By adopting SVM with the selected LBPH bins, we obtain the classication rate of 94.81% on the LFW database. We obtain the performance of 96.79% by applying boosting learning on LFW dataset 92.04% by applying Support Vector Machine (SVM) on RUS-FD dataset. The approach proposed in this paper is promising to be further studied on other face classification tasks, such as age estimation and emotion recognition.

ACKNOWLEDGMENT

This work was supported by Russian Foundation for Basic Research grants (№ 15-07-08674 and № 15-08-99639).

REFERENCES

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

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

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

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

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

[6] J. Maydt and R. Lienhart, "Face Detection with Support Vector Machines and a Very Large Set of Linear Features", *in IEEE ICME 2002,* Lousanne, Switzerland, 2002.

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

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

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

[10] 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", *in IEEE Int. Conf. on Control Automation Robotics & Vision (ICARCV)*, 2010, pp. 1280-1284.

[11] K.K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection", *in IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, 1998, pp. 39-51.

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

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

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

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

[16] A. B. A. Graf and S. Borer, "Normalization in Support Vector Machines. Proceedings of the DAGM", *LNCS 2191*, 2001.

[17] L.K. Saul and S. T. Roweis, "An Introduction to Locally Linear Embedding", *in Report at AT&T Labs - Research*, 2000.

[18] B. Moghaddam and M.H. Yang, "Learning gender with support faces", *in Trans. Pattern Anal. Mach. Intell.*, vol. 24(5), 2002, pp. 707– 711.

[19] M. Toews and T. Arbel, "Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion", *in Trans. Pattern Anal. Mach. Intell.*, vol. 31(9), 2009, pp. 1567–1581.

[20] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces", *in In Proc. Conf. Comput. Vision Pattern Recognition*, vol. 1, 2006, pp. 387–394.

[21] T. Wallace and P. Wintz, "An efficient three-dimensional aircraft recognition algorithm using normalized fourier descriptors", *in Computer Vision Graphics Image Processing*, 1980, pp. 13.

[22] G. Shakhnarovich, P. Viola, and B. Moghaddam, "A Unified Learning Framework for Real Time Face Detection & Classification", *in Int Conf. on Automatic Face & Gesture Recognition – FG 2002*, May 2002, pp. 16 – 26.

[23] B. Wu, H. Ai, C. Huang and S. Lao, "Fast rotation invariant multi-view face detection based on real Adaboost", *6th Int. Conf. on Face and Gesture Recognition - FG 2004,* Seoul, Korea, May 2004, pp. 79–84.

[24] S. Gutta, H. Wechsler and P. J. Phillips, "Gender and ethnic classification", *in Proceedings of the IEEE International Automatic Face and Gesture Recognition,* 1998, pp. 194–199.

[25] Z. Sun, G. Bebis, Xiaoping Yuan and Sushi J. Louis, "Genetic Feature Subset Selection for Gender Classification", *in A Comparison Study, IEEE Workshop on Applications of Computer Vision,* 2002, pp.165-170.

[26] G. Shakhnarovich, P.A. Viola, and B. Moghaddam, "A Unified Learning Framework for Real Time Face Detection and Classification", *in Proc. Int'l Conf. Automatic Face and Gesture Recognition*, 2002, pp. 14-21.

[27] R. Brunelli and T. Poggio, "Hyperbf networks for gender classification", *in DARPA Image Understanding Workshop*, 1992, pp: 311–314.

[28] Viola P., Jones M. J., "Robust Real-Time Face Detection", *in International Journal of Computer Vision*, vol. 57(2), 2004, pp. 137-154.

[29] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and Gang Hua, "Labeled Faces in the Wild: A Survey", *in Advances in Face Detection and Facial Image Analysis*, Springer, 2016, pp. 189-248.

[30] C. Shan, "Learning Local Binary Patterns for Gender Classification on Real-World Face Images", *in Pattern Recognition Letters*, vol. 33 issue 4, March 2012, pp. 431-437.