The Development of Object Tracking and Recognition Algorithms for Audience Analysis System

Vladimir Khryashchev, *Member, IAENG*, Lev Shmaglit, Maxim Golubev and Andrey Shemyakov

***Abstract*—A system of video data analysis based on computer vision methods is presented. The proposed system consists of four consecutive stages: face detection, face tracking, gender recognition and age classification. The proposed software complex can find its applications in different areas, from digital signage and video surveillance to automatic systems of accident prevention and intelligent human-computer interfaces.**

***Index Terms*—Image recognition, face detection, gender classification, age estimation, machine learning, object tracking, support vector machines**

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]. 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. [2]. 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.

Manuscript received May 3, 2013, revised May 18, 2013.

Vladimir Khryashchev is with the Image Processing Laboratory, P.G. Demidov Yaroslavl State University, 14 Sovetskaya, Yaroslavl, Russia, 150000 (phone: +74852797775; e-mail: vhr@yandex.ru).

Lev Shmaglit is with the Image Processing Laboratory, P.G. Demidov Yaroslavl State University, 14 Sovetskaya, Yaroslavl, Russia, 150000 (phone: +74852797775; e-mail: Shmaglit_Lev@yahoo.com).

Maxim Golubev is with the Image Processing Laboratory, P.G. Demidov Yaroslavl State University, 14 Sovetskaya, Yaroslavl, Russia, 150000 (phone: +74852797775; e-mail: maksimgolubev@yandex.ru).

Andrey Shemyakov is with the Image Processing Laboratory, P.G. Demidov Yaroslavl State University, 14 Sovetskaya, Yaroslavl, Russia, 150000 (e-mail: andrey.shemiakov@gmail.com).

Gender recognition, for example, can be used to collect and estimate demographic indicators [3-6]. 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.

Human age estimation is another problem in the field of computer vision which is connected with face area analysis [7]. Among its possible applications one should note electronic customer relationship management (such systems assume the usage of interactive electronic tools for automatic collection of age information of potential consumers in order to provide individual advertising and services to clients of various age groups), security control and surveillance monitoring (for example, an age estimation system can warn or stop underage drinkers from entering bars or wine shops, prevent minors from purchasing tobacco products from vending machines, etc.), biometrics (when age estimation is used as a part that provides ancillary information of the users' identity information, and thus decreases the whole system identification error rate). Besides, age estimation can be applied in the field of entertainment, for example, to sort images into several age groups, or to build an age-specific human-computer interaction system, etc. [7].

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. The block diagram of the proposed system is presented in fig. 1.

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 [8], 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

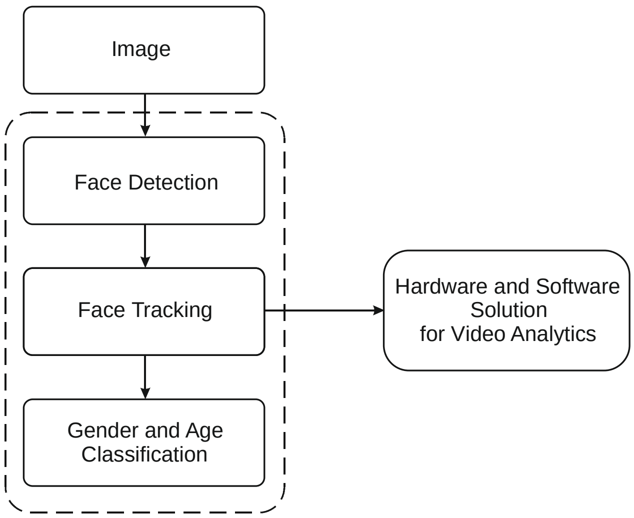


Fig. 1. A block diagram of the proposed audience gender and age classification system.

age groups.

To estimate the period of a person’s stay in the range of camera’s visibility, face tracking algorithm is used. Generally speaking, the task of tracking is to match same objects on different frames of a video sequence. Object tracking itself is a difficult problem, as it is influenced simultaneously by the following factors:

– the variation of image parameters, scene illumination and camera noise;

– the presence of objects with varying form (for example, the running person);

– temporary disappearance of analyzed objects due to overlapping by other objects;

– the existence of several moving objects at the same time with similar features and crossed trajectories;

– distortions due to wrong segmentation of objects at the previous processing stages.

The main approach to age estimation is a two-level scheme where on the first step special features [9] are extracted from the analyzed image fragment (the best results are reached applying the combination of various feature descriptors, such as AAM, HOG, LBP, DSIFT, etc.), and on the second step a classifier is used to find areas in the resulted feature space, corresponding to certain ages (directly, or by means of a set of binary classifiers and a voting scheme). The best results of classification can be reached by utilizing a combination of various approaches, such as SVM (Support Vector Machines), ANN (Artificial Neural Networks), RF (Random Forests), etc. [7]. Difficult hierarchical schemes, applied to classifier design, also allow to achieve an advantage in some cases [10].

The study of age estimation performance under variations across race and gender [11] discovered that crossing race and gender can result in significant error increase. Age facial features of men and women, and also of representatives of various races significantly differ from each other. Thus the optimum strategy of training is to form a number of separate training sets and to construct an independent classifier for each analyzed category. Age estimation is then conducted with preliminary division of all input images into defined categories in order to choose a suitable classifier for each image. The problem of such scheme with preliminary division into categories lies not only in the increase of computational complexity of the total classifier but, mainly, in the significant increase of the required training database capacity.

In paper [12] the framework, described above, is utilized for age estimation of faces varying by their relative position to camera (frontal, panning in horizontal and vertical direction). In work [13] a different strategy is suggested to improve the accuracy of age estimation under facial expression changes. It lies in the search of correlation between faces with different facial expression and in the conversion of initial feature space into space where features are similar (become invariant) for neutral, smiling and faces with a sad expression.

In paper [14] automatic facial alignment based on eye detection is suggested to reduce the influence of face position variation.

The rest of the paper briefly describes main algorithmic techniques utilized on different stages of the proposed system. The level of gender and age classification accuracy is estimated in real-life situations. It should be also noted that algorithms, proposed in this paper, incorporate universal machine learning techniques, and thus can be applied to solve any other problems of object classification and image understanding.

II. Face Detection

To solve the problem of face detection an algorithm, suggested by P. Viola and M. Jones in paper [8], was chosen. It utilizes a learning procedure based on adaptive boosting [15-18]. This procedure consists of three parts:

1) **Integral image representation.**

The integral image at location (x, y) contains the sum of the pixels above and to the left of x, y, inclusive:

$$ii(x,y) = \sum_{x' \le x,\, y' \le y} i(x',y'),$$

where $ii(x,y)$ is the integral image, $i(x,y)$ is the original image. Integral image representation allows to speed up the calculation of a rectangular feature set as any rectangular sum can be computed in four array references.

2) **Learning classification functions using AdaBoost.**

For each feature, the weak learner determines the optimal threshold classification function, such that the minimum number of examples are misclassified. A weak classifier ($h(x, f, p, \theta)$) thus consists of a feature (f), a threshold (θ) and a polarity (p) indicating the direction of the inequality:

$$h_j(x) = \begin{cases} 1, & \text{if } \; p_j f_j(x) < p_j \theta_j \\ 0, & \text{otherwise} \end{cases},$$

where x is a 24 × 24 pixel sub-window of an image.

3) **Combining classifiers in a cascade structure**, which allows background regions of the image to be quickly discarded while spending more computation on promising face-like regions. The cascade structure of a resulted classifier is schematically presented in fig. 2. It consists of N layers each of which represents a classifier generated by the

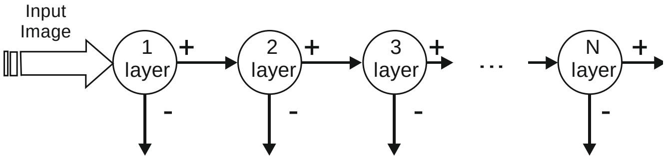


Fig. 2. Schematic depiction of the detection cascade.

AdaBoost learning procedure.

The considered algorithm is one of the most widely used to solve the problem of face detection on digital images. It is a part of computer vision library OpenCV (Open Source Computer Vision Library) [18].

III. Face Tracking

Nowadays there exist several approaches to the realization of object tracking on video sequences. For real time applications methods based on the estimation of optical flow are most widely used.

Generally speaking, optical flow can be defined as a two dimensional projection of objects motion on an image plane, representing object pixel's trajectories. Optical flow can be calculated as on the basis of tracking of all image pixels (full optical flow) and on the basis of tracking of particular feature pixels (sparse optical flow) [2]. The disadvantage of full optical flow is its low resistance to image distortions caused by the presence of noise. Thus, the calculation of sparse optical flow is used more often in practice.

An algorithm, proposed by B. Lucas and T. Kanade in paper [19], was chosen as the basic approach to solve the problem of optical flow calculation. With the help of this algorithm the coordinates of feature pixels on the current image frame are calculated out of their coordinates on the previous frame. Then the estimation of new position and size of a tracked object is performed on the basis of the following original algorithm (fig. 3).

1) **Object offset calculation.** First, the differenced in coordinates between features on the previous frame and current features are calculated. The resulted values are processed by median filter in order to smooth the spikes, obtained due to inaccuracies in tracking of some certain pixels. Such inaccuracies may be, for example, caused by the presence of noise. After that filtered offsets of feature pixels are averaged. The obtained average value is used as a total offset of the whole object.

2) **Scaling coefficient calculation.** Due to the fact that the object is scaled relative to its center, for scaling coefficient calculation the coordinates of all feature pixels are recalculated relative to the center of tracked object. Then the distance from each feature to the center of the object is calculated. The resulted values are processed by median filter and then averaged. The obtained average value is rounded to the second sign and used as a scaling coefficient of tracked object.

After the estimation of position and size of a tracked object on the current frame, feature pixels, which lie outside the defined border, are rejected.

Face window scaling and offset calculation is not the only problem which Lucas-Kanade method does not solve. The other two are face overlapping and face crossing. We propose three modifications to solve the problems described above. The first one (Lucas-Kanade-1) defines tracking object offset and scaling factor as the averages of key pixels corresponding characteristics. The second one (Lucas-Kanade-2) introduces median filtration to improve the overall performance. The third one (Lucas-Kanade-3) detects overlapping and crossing by dividing the window

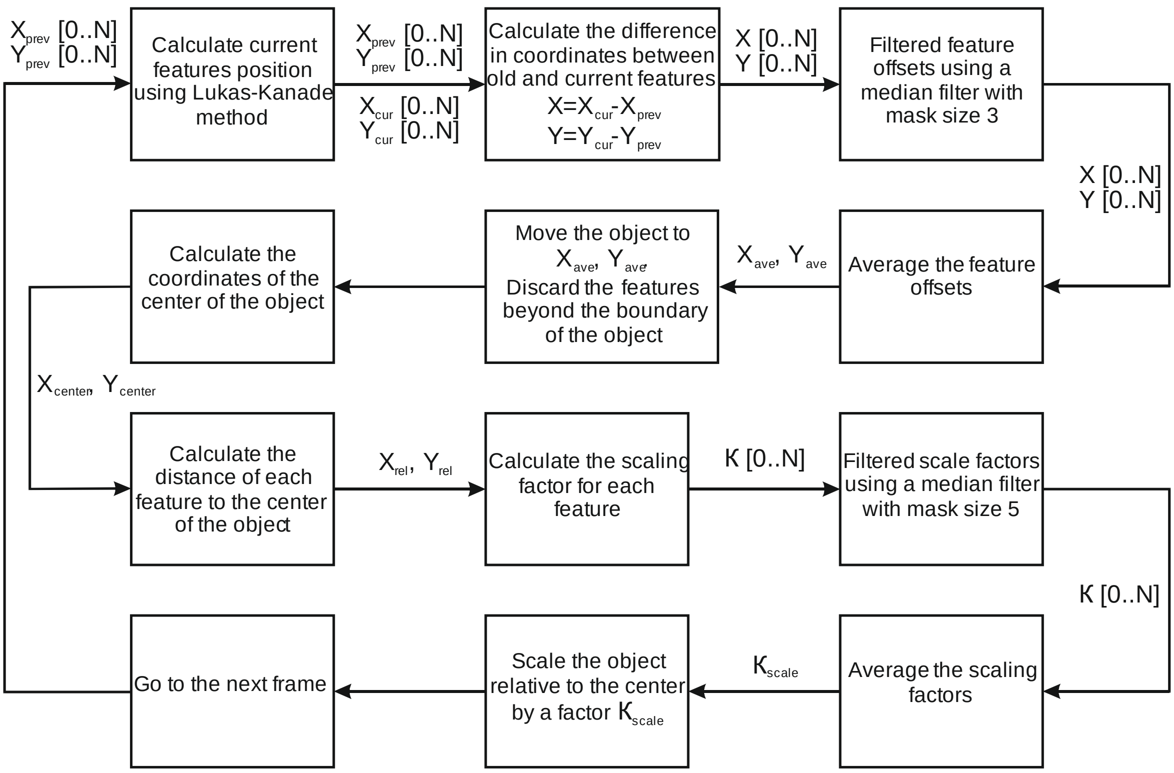


Fig. 3. The scheme of modified tracking algorithm.

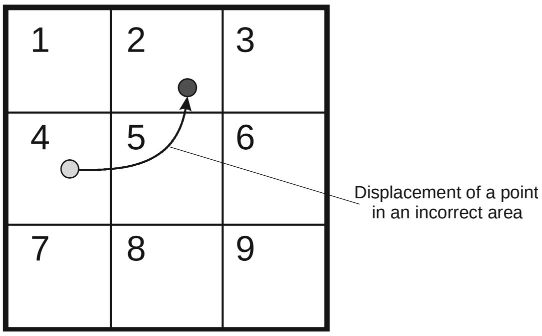


Fig. 4. Overlapping and crossing pixel detection.

into square regions and labelling each key pixel to the corresponding region. If a pixel moves out of its region, as it is showed in fig. 4, then such pixel is removed from further consideration being suspected as an overlapped one.

In order to compare the proposed modifications, a test sequence containing very difficult movement was chosen (fig. 5). It comprises all necessary tracking difficulties including overlapping, fast movement and camera trembling.

In fig. 6 the performance of the proposed Lukas-Kanade modifications on the chosen test video sequence is presented. We define tracking rate as a relation of the number of frames on which tracking window follows the object to the total number of frames in a video segment.

The results show that the proposed second and third modifications allow to achieve better tracking rate compared to classic Lucas-Kanade-1. Lucas-Kanade-2 misses the object only once due to overlapping, while Lucas-Kanade-3 does not make mistakes in such conditions but misses the object at the very end of test sequence due to camera trembling. Thus, we can make a conclusion that the proposed face tracking technique requires further improvement.

Fig. 5. Test video sequence.

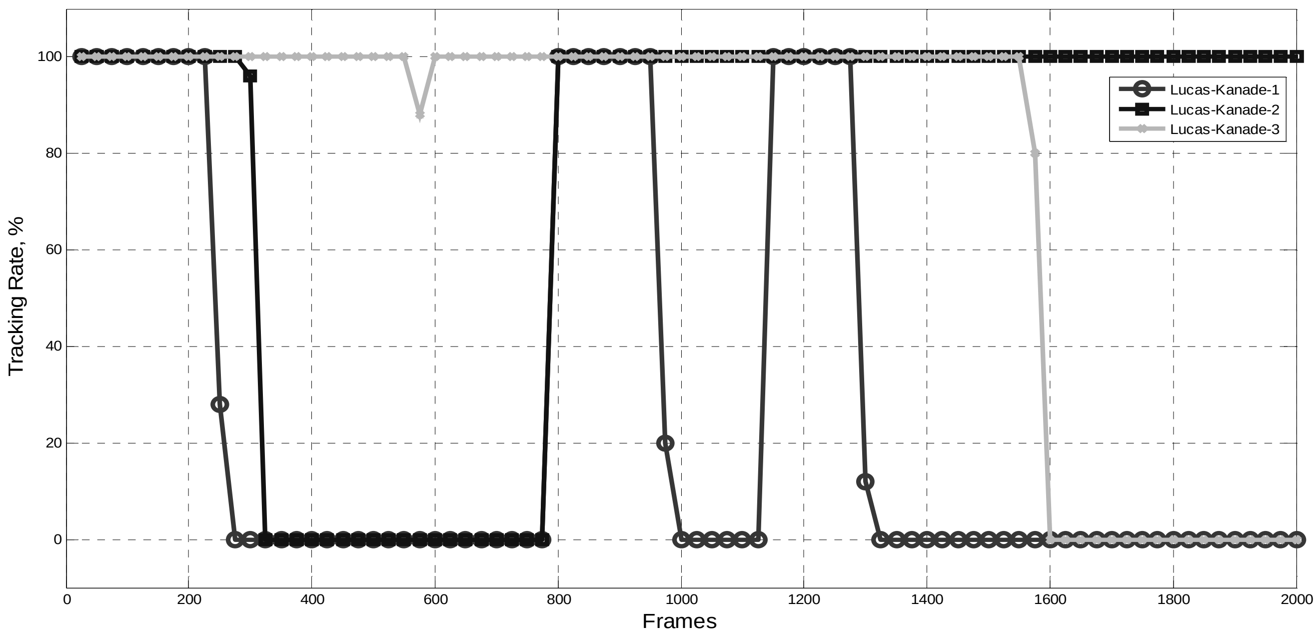


Fig. 6. Tracking algorithms testing results.

IV. Gender Recognition

A new 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. In order to construct a fully automatic face analysis system, gender recognition is used in connection with AdaBoost face detection classifier, described in section 2. Detected fragments are preprocessed to align their luminance characteristics and to transform them to uniform scale.

Classifier is based on Adaptive Features and SVM (AF-SVM). Its operation includes several stages, as shown in fig. 7. 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.

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, 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.

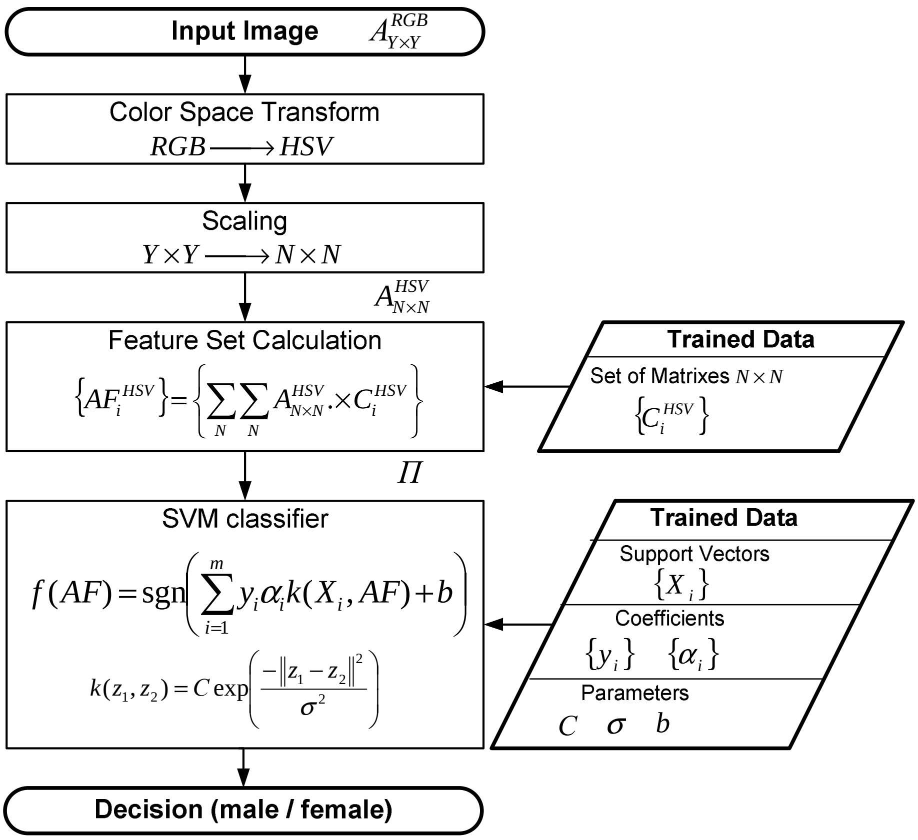


Fig. 7. The scheme of the proposed gender classification algorithm.

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 [7]:

$$F = \frac{\left(\left\langle \left\{AF_i^{HSV}\right\}_M \right\rangle - \left\langle \left\{AF_i^{HSV}\right\}_F \right\rangle\right)^2}{\sigma\left\{AF_i^{HSV}\right\}_M + \sigma\left\{AF_i^{HSV}\right\}_F};$$

– iteratively in a cycle (until the number of iterations exceeds some preliminary fixed maximum value): random generation of coefficient matrix $\tilde{C}_i^{HSV}$ inside the fixed neighborhood of matrix C_i^{HSV}, feature value $\tilde{AF}_i^{HSV}$ calculation for each training fragment, calculation of the utility function $\tilde{F}$, transition to a new point ($F \to \tilde{F}, C \to \tilde{C}$), if $\tilde{F} > F$;
– 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. 8. Maximum recognition rate (about 80%) was obtained for $C = 10^6$ and $\sigma = 10^8$.

Besides, we investigated the dependence of classifier performance from the number of features extracted from each color component – N2, and from the number of training images for each class – M2 (fig. 9).

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

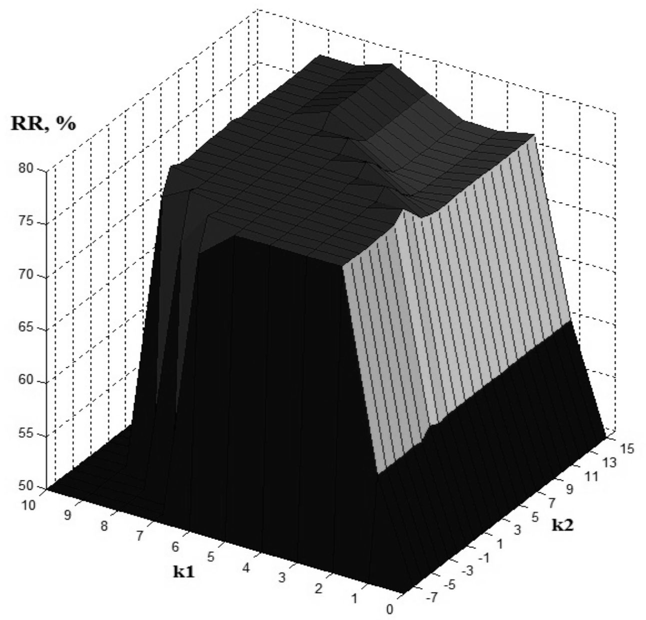


Fig. 8. The dependence of RR from kernel function parameters $C = 10^{k1}$ and $\sigma = 10^{k2}$.

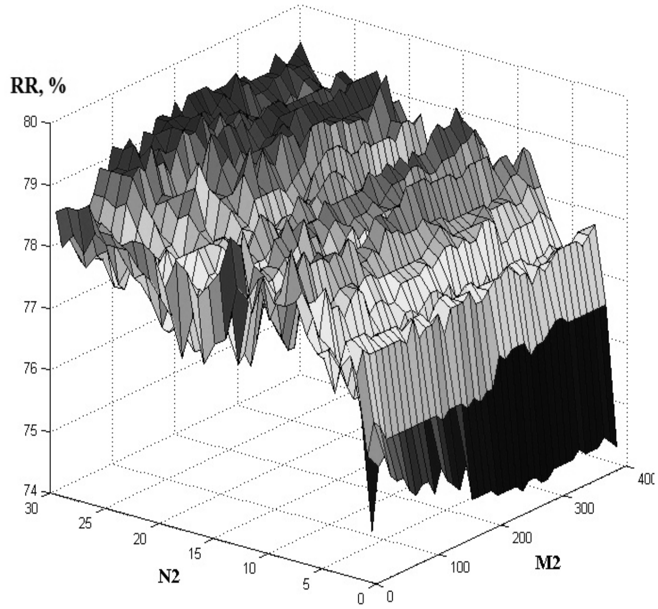


Fig. 9. The dependence of recognition rate from training procedure parameters.

following parameters were chosen: $N2 = 30$; $M2 = 400$.

Let's turn to the results of the proposed AF-SVM algorithm comparison with state-of-the-art classifiers: SVM and KDDA (Kernel Direct Discriminant Analysis).

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). 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.

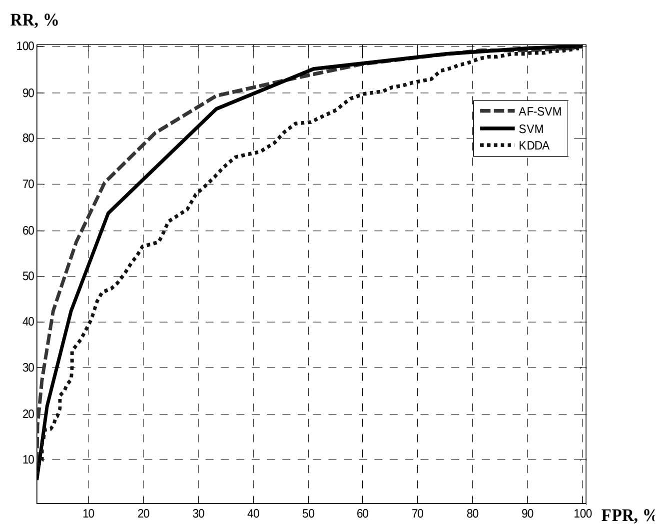


Fig. 10. ROC-curves of tested gender recognition algorithms.

The results of AF-SVM, SVM and KDDA testing are presented in fig. 10 and in table 2. 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 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 utilize 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 (containing 400 training images per class) from the whole training database (containing 5000 images per class). Thus

TABLE II
COMPARATIVE ANALYSIS OF TESTED ALGORITHMS PERFORMANCE

Parameter \ Algorithm	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	

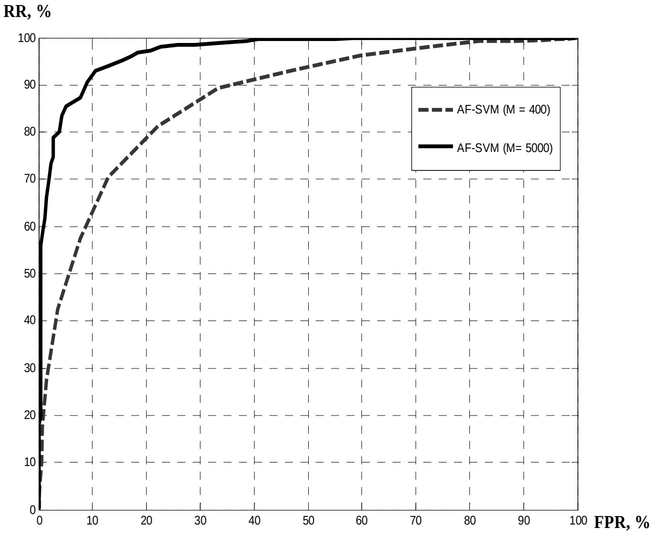


Fig. 11. ROC-curves for AF-SVM algorithm trained on datasets of different size.

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.

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 3 and in fig. 11.

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

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

nearly 91% is achieved. It should be also noted that 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).

V. AGE ESTIMATION

The proposed age estimation algorithm realizes hierarchical approach (fig. 12). First of all input fragments are divided into three age groups: less than 18 years old, from 18 to 45 years old and more than 45 years old. After that the results of classification on the first stage are further divided into seven new groups each of which is limited to one decade. Thus the problem of multiclass classification is reduced to a set of binary “one-against-all” classifiers (BC). These classifiers calculate ranks for each of the analyzed classes. The total decision is then obtained by the analysis of the previously received histogram of ranks.

A two level scheme of binary classifier construction is applied with the transition to adaptive feature space, similar to described earlier, and support vector machines classification with RBF kernel.

Input fragments are preprocessed to align their luminance characteristics and to transform them to uniform scale. Preprocessing includes color space transformation and scaling, both similar to that of gender recognition algorithm. Features, calculated for each color component, are combined

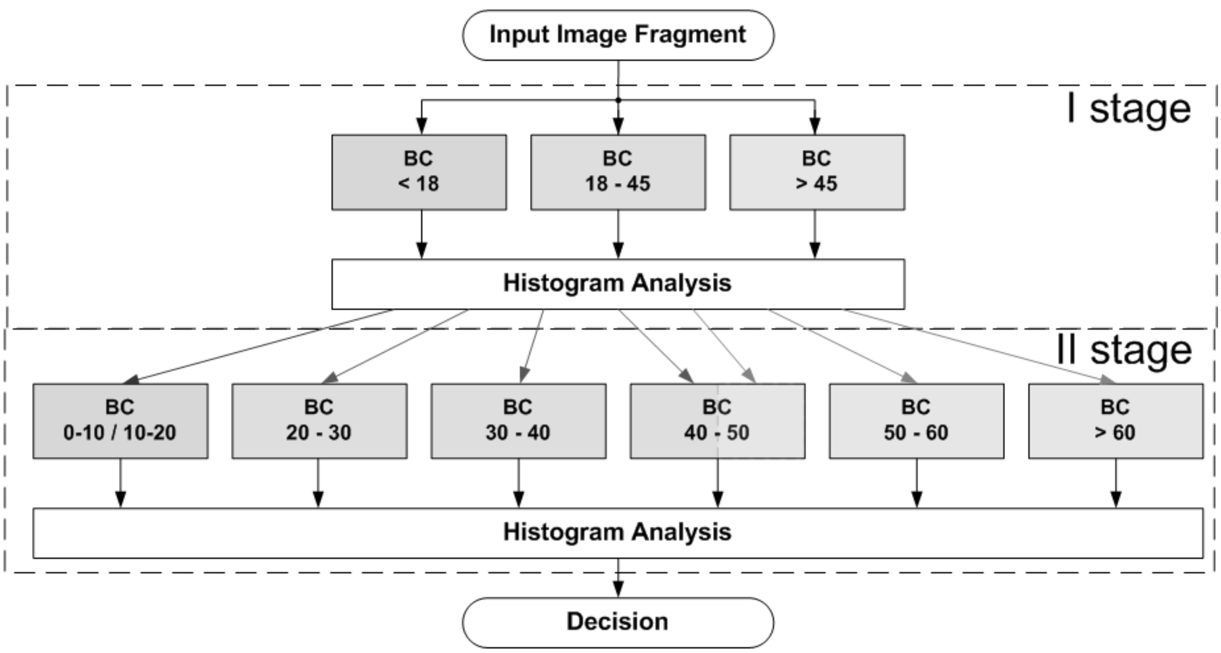


Fig. 12. A block diagram of the proposed age estimation algorithm.

to form a uniform feature vector.

Training and testing require a huge enough color image database. We used state-of-the-art image databases MORPH [20] and FG-NET [21] and our own image database, gathered from different sources, which consisted of 10 500 face images. Faces on the images were detected automatically by AdaBoost face detection algorithm.

A total number of seven thousand images were used for age classification algorithm training and testing on the first stage (table 4). 3 binary classifiers were constructed utilizing 144 adaptive features each (table 5).

Their ROC Curves are presented in fig. 13. It is clear, that

TABLE IV
FIRST STAGE IMAGE DATABASE PARAMETERS

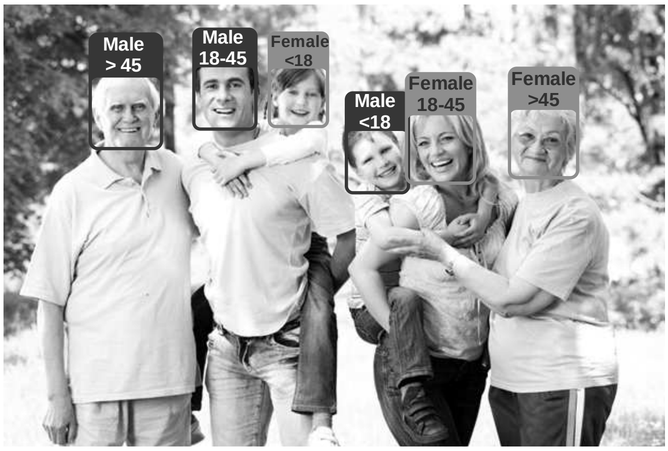
Class Label / Database Capacity	<18	18-45	>45	Total
Training Images per Class	2000	2000	3000	7000
Testing Images per Class	226	400	531	1157
Total Number of Images	2226	2400	3531	8157

TABLE V
FIRST STAGE AF-SVM SUMMARY

Training Parameters		Value
Number of Binary Classifiers		3
Number of Color Components Used		3
Number if Adaptive Features Generated	Per Color Component	48
	Per Binary Classifier	144
	Total	432

TABLE VI
FIRST STAGE AGE CLASSIFICATION RESULTS

Decision / Ground Truth	<18	18-45	>45
<18	**82%**	14%	4%
18-45	22%	**58%**	20%
>45	3%	5%	**92%**



a)

b)

Fig. 14. Visualization of the proposed system age estimation performance: on the first stage (a); on the second stage (b).

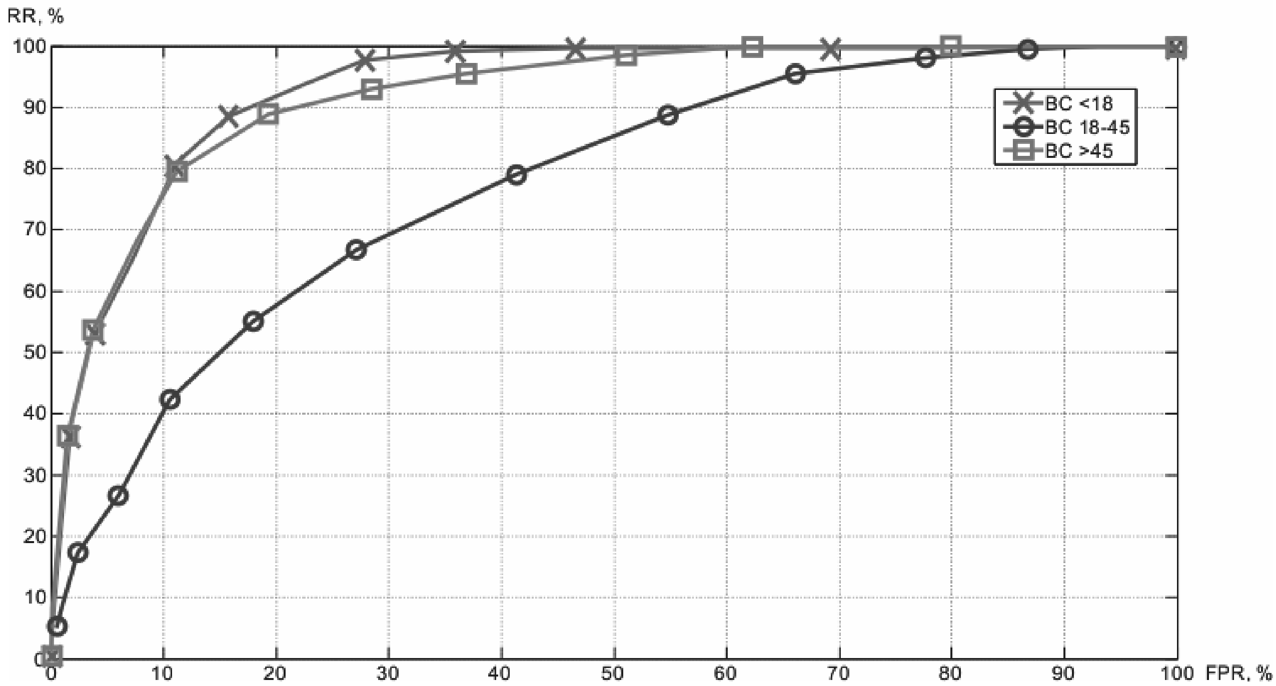


Fig. 13. Binary classifiers ROC-curves.

the main problem is to distinguish an age group from 18 to 45 years old.

Classification results are as follows (table 6): 82% accuracy for young age group, 58% accuracy for middle age group and 92% accuracy for senior age group. Age classification rate in a three age group division problem – 77.3%.

Binary classifiers of the second stage were constructed similar to those of the first stage described above. A visual example of age estimation by the proposed algorithm on its first and second stages is presented in fig. 14(a) and 14(b) respectively.

VI. Conclusion

The system, described in this paper, provides collection and processing of information about the audience in real time. It is fully automatic and does not require people to conduct it. No personal information is saved during the process of operation. A modern efficient classification algorithm allows to recognize viewer's gender with more than 90% accuracy.

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.

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] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

[18] Open Source Computer Vision Library, http://sourceforge.net/projects/opencvlibrary/.

[19] B. Lucas, T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proceedings of Imaging Understanding Workshop*, pp. 121-130, 1981.

[20] K. Ricanek, T. Tesafaye, "MORPH: A Longitudinal Image Database of Normal Adult Age-Progression," *IEEE 7th Int. Conf. on Automatic Face and Gesture Recognition*, pp. 341-345, 2006.

[21] The FG-NET Aging Database, http://www.fgnet.rsunit.com/, http://wwwprima.inrialpes.fr/FGnet/, 2010.