Audience Analysis System on the Basis of Face Detection, Tracking and Classification Techniques

Vladimir Khryashchev, *Member, IAENG*, Alexander Ganin, Maxim Golubev, and Lev Shmaglit

***Abstract*—A system of video data analysis based on computer vision methods is presented. The main algorithms of the proposed software complex, that are face detection, face tracking, gender recognition and age estimation, are described. The proposed classifier on the basis of adaptive features and support vector machines allows to increase the accuracy of gender and age recognition.**

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

I. Introduction

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. [1, 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; automatic systems of accident prevention, intelligent human-computer interfaces.

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.

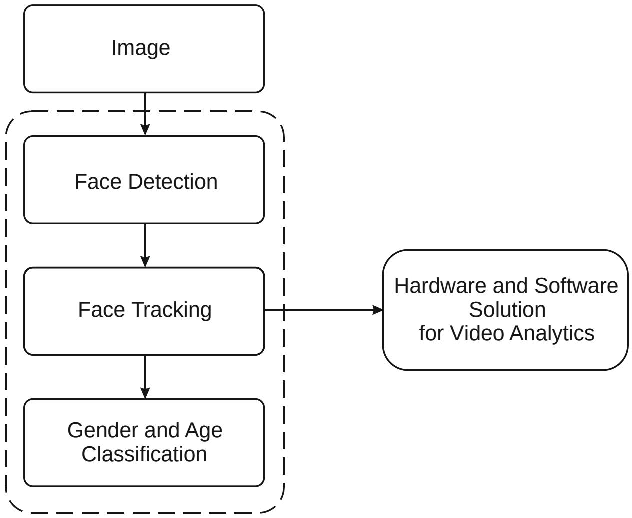


Fig. 1. A block diagram of the proposed audience gender and age classification system.

This work was supported in part by the Russian Foundation for Basic Research under Grant № 12-08-01215-a "Development of methods for quality assessment of video".

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

Alexander Ganin, PhD Student, Image Processing Laboratory, P.G. Demidov Yaroslavl State University, 14 Sovetskaya, Yaroslavl, Russia, 150000 (phone: +74852797775; e-mail: connect@piclab.ru).

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

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

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 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 [15-17]. 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:

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

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 AdaBoost learning procedure.

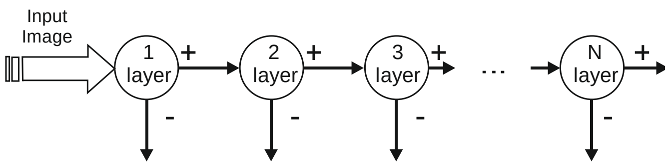


Fig. 2. Schematic depiction of the detection cascade.

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

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.

IV. Gender Recognition

A 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. Thus, the proposed classifier is based on Adaptive Features and SVM (AF-SVM) [6].

AF-SVM algorithm consists of the following steps: color space transform, image scaling, adaptive feature set calculation and SVM classification with preliminary kernel transformation. Data, required for the calculation of adaptive feature set, is generated during training. The training procedure of the proposed AF-SVM classifier can be split into two independent parts:

– feature generation,
– SVM construction and optimization.

Each stage of training includes a set of adjustable

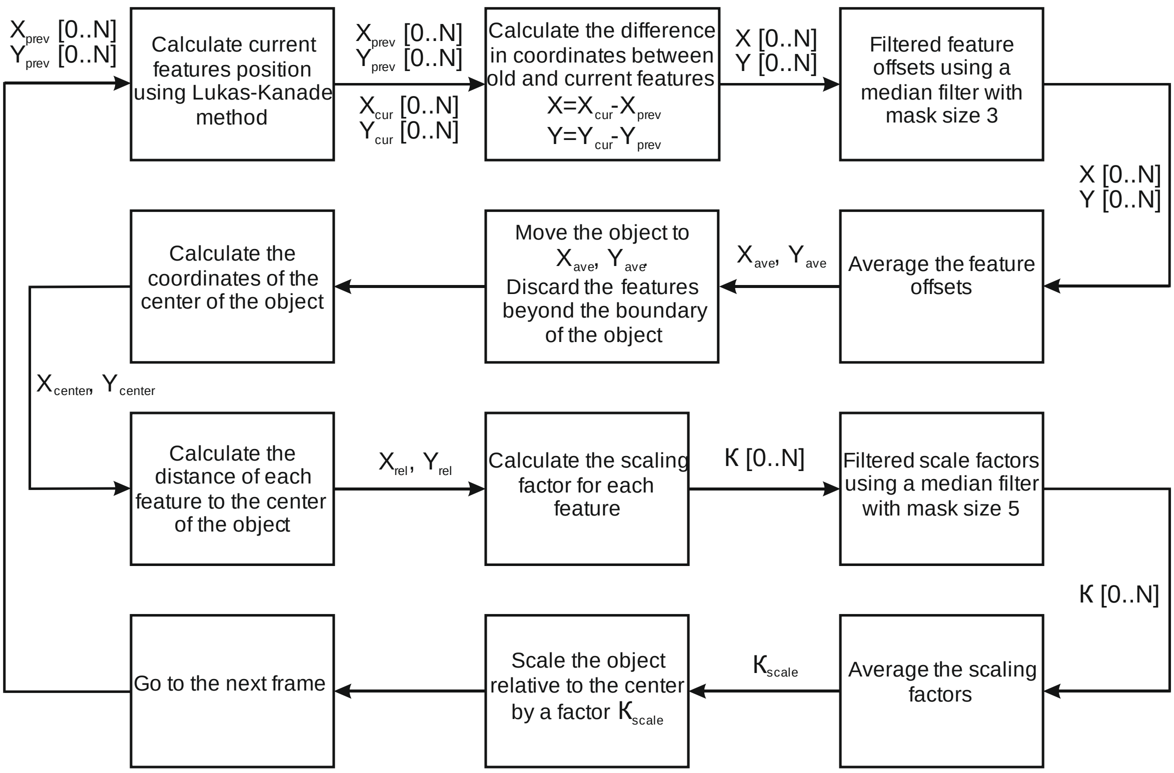


Fig. 3. The scheme of modified tracking algorithm.

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

parameters which affect the resulted efficiency of the classifier. These parameters optimization is an essential part of the training procedure [6].

In table 1 we present the results of the proposed AF-SVM algorithm comparison with two state-of-the-art classifiers: classic SVM and KDDA (Kernel Direct Discriminant Analysis).

TABLE I
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	90.6	9.4
Classified as "female", %	75.5	24.5	65.5	34.5	91	9
Total classification rate, %	**77.7**	22.3	**69.7**	30.3	**90.8**	9.2
Operation speed, faces/sec	44		45		65	

SVM and KDDA are one of the most effective training techniques applied to the task of binary classification of digital images. 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.

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 – 90.8% 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.

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. 4). 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 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

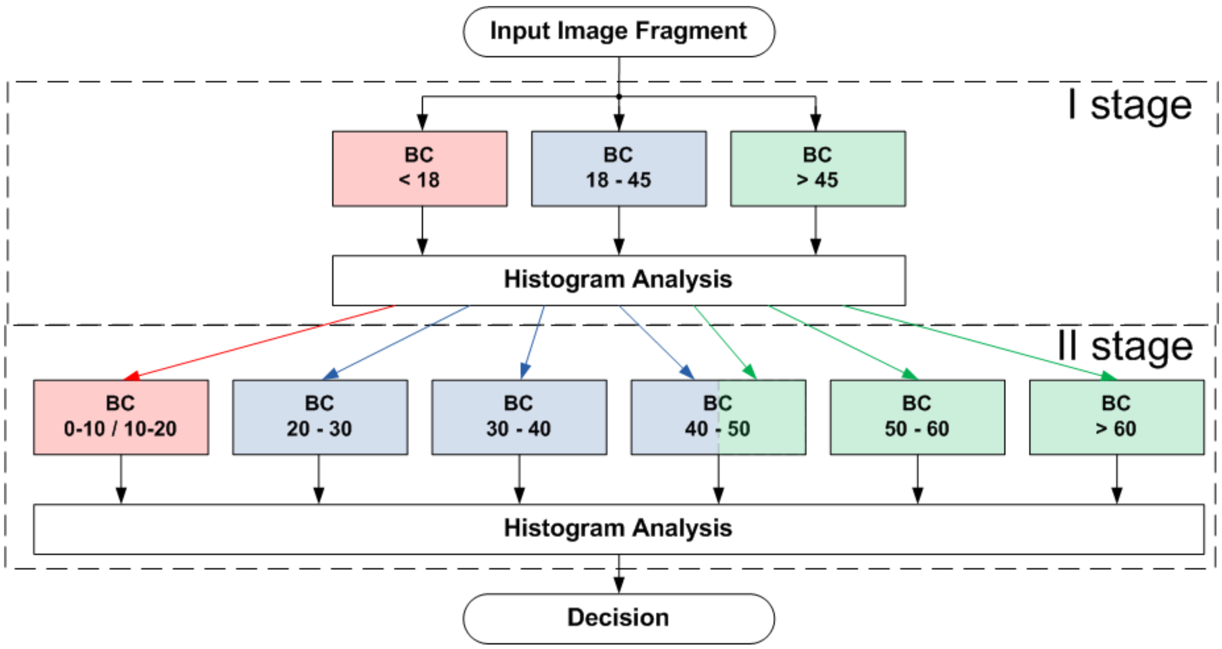


Fig. 4. A block diagram of the proposed age estimation algorithm.

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

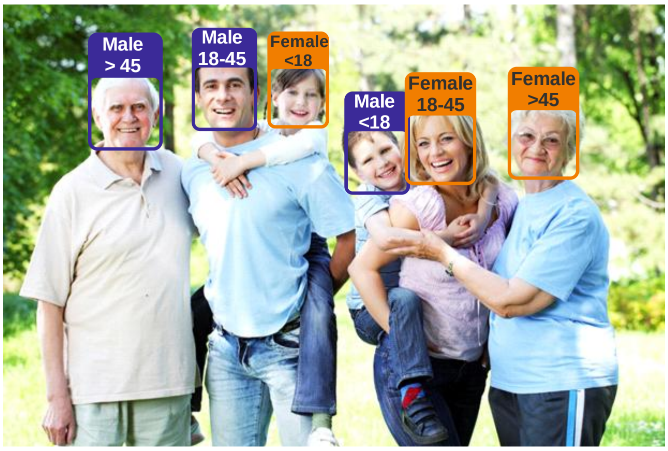
face images. Faces on the images were detected automatically by AdaBoost face detection algorithm.

A visual example of age estimation by the proposed algorithm on its first and second stages is presented in fig. 5(a) and 5(b) respectively.

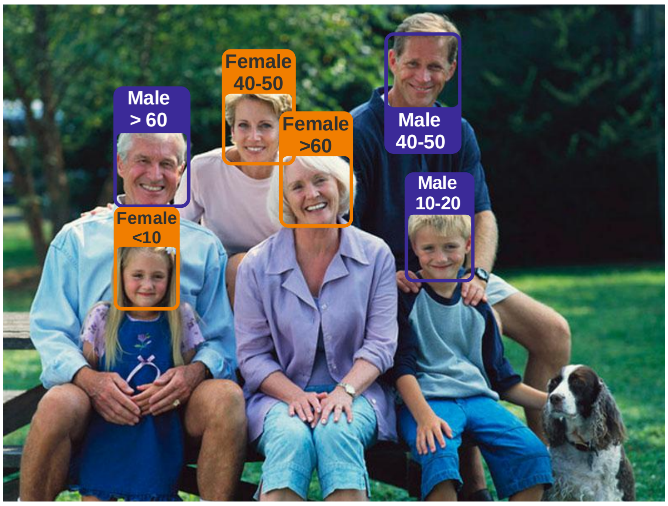
VI. Conclusion

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.



a)



b)

Fig. 5. Visualization of the proposed system age estimation performance: on the first stage (a); on the second stage (b).

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] L. Shmaglit, M. Golubev, A. Ganin, V. Khryashchev, "Gender classification by face image," *Proc. Int. Conf. on Digital Signal Processing and its Application (DSPA)*, vol. 1, pp. 425-428, 2012.

[6] 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.

[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.

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)