# Online Audience Measurement System Based on Machine Learning Techniques

Vladimir Khryashchev[(✉)], Andrey Priorov, and Alexander Ganin

P.G. Demidov Yaroslavl State University, Yaroslavl, Russia
dcslab@uniyar.ac.ru

**Abstract.** An application for video data analysis based on computer vision methods is presented. The proposed system consists of five consecutive stages: face detection, face tracking, gender recognition, age classification and statistics analysis. AdaBoost classifier is utilized for face detection. A modification of Lucas and Kanade algorithm is introduced on the stage of tracking. Novel gender and age classifiers based on adaptive features, local binary patterns and support vector machines are proposed. More than 92 % accuracy of viewer's gender recognition is achieved. All the stages are united into a single system of audience analysis. The system allows to extract all the possible information about depicted people from the input video stream, to aggregate and analyze this information in order to measure different statistical parameters.

**Keywords:** Online audience measurement · Machine learning · Gender classification · Age estimation · Local binary patterns · Support vector machines

## 1 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 [1–6]. 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. [7]. 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 [8–10]. 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 [11]. 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. [11].

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 [12–17]. In this paper we propose 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).
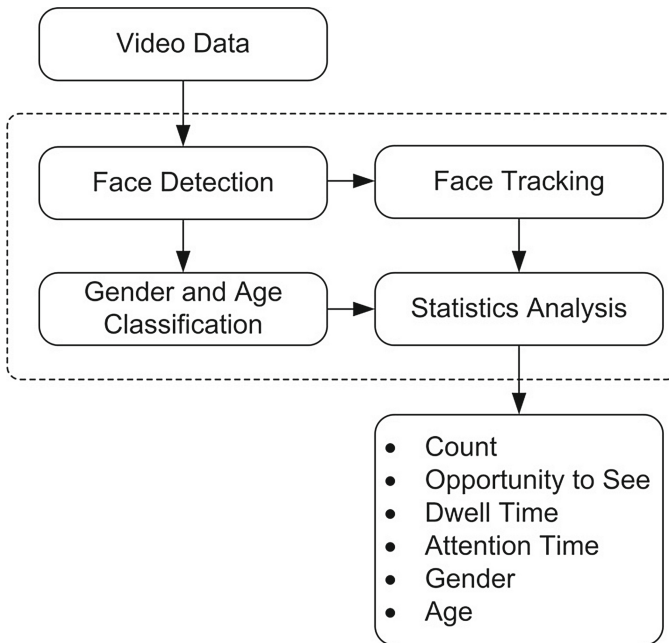


**Fig. 1.** A block diagram of the proposed application for video analysis

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 [18], 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. The proposed gender and age classifiers are based on non-linear SVM (Support Vector Machines) classifier with RBF kernel. To extract information from image fragment and to move to a lower dimension feature space LBP features are utilized.

To estimate the period of a person's stay in the range of camera's visibility, face tracking [19–22] algorithm is used. It is based on Lucas-Kanade optical flow calculation procedure [23].

The rest of the paper briefly describes main algorithmic techniques utilized on the stages of gender and age recognition. The level of gender and age classification accuracy is estimated in real-life situations.

## 2 Gender Recognition

A new gender recognition algorithm, proposed in this paper, is based on non-linear SVM classifier with RBF kernel. Detected fragments are preprocessed to align their luminance characteristics and to transform them to uniform scale. After that to extract information from image fragment and to move to a lower dimension feature space local binary patterns (LBP) [24] operator is utilized. These simple local features have been proved to show good results in application to face recognition tasks. Their calculation procedure is shown in Fig. 2.
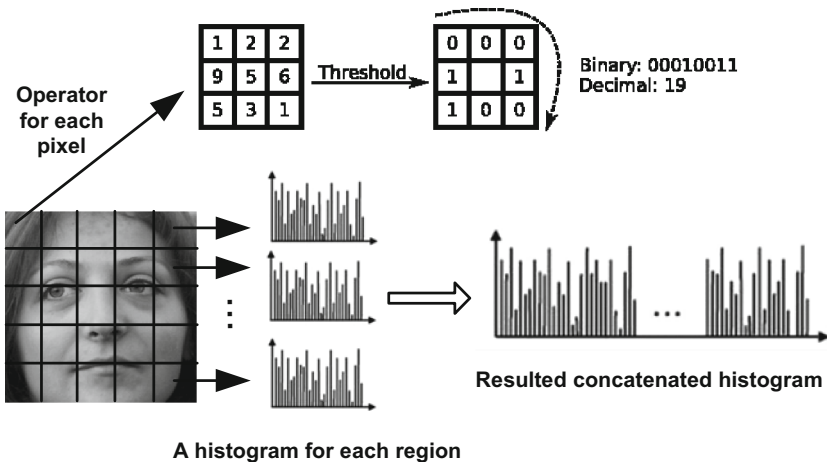


**Fig. 2.** LBP feature vector extraction procedure

On the first step each pixel is compared with its neighbors. The result of comparison is presented in binary scale. These digits from a given neighborhood (let's say $3 \times 3$ pixels) form a binary number which can be presented in decimal format.

On the second stage image is divided into rectangular regions. A histogram of frequencies of emergence of numbers, acquired on the first step, is calculated for each region. The resulted feature vector is a concatenation of histograms from all regions.

The obtained feature vector is transformed using a Gaussian radial basis function kernel using Eq. 1:

$$k(z_1, z_2) = C \exp\left(\frac{-\|z_1, z_2\|^2}{\sigma^2}\right) \tag{1}$$

Kernel function parameters $C$ and $\sigma$ are defined during training. The resulted feature vector serves as an input to linear SVM classifier which decision rule is specified by Eq. 2:

$$f(AF) = \text{sgn}\left(\sum_{i=1}^{m} y_i \alpha_i k(X_i, AF) + b\right). \tag{2}$$

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. This is how the proposed gender classifier based on LBP features and SVM was constructed (LBP-SVM classifier).

Both gender recognition algorithm training and testing require big enough color image database. The most commonly used image database for the tasks of human faces recognition is the FERET database [25], 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 and Fig. 3).

**Table 1.** 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 |

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

a) male faces   b) female faces

**Fig. 3.** Detected fragments from the proposed image database

obtained. This dataset was split into three independent image sets: training, validation and testing. Training set was utilized for 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.

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.

The proposed classifier was compared to AF-SVM algorithm described in paper [10]. AF-SVM was chosen as a reference because it has both high recognition rate and low operational complexity compared to state-of-the-art classifiers [26].

Testing results of the proposed LBP-SVM classifier compared to AF-SVM performance are presented in Table 2 and Fig. 4.

**Table 2.** Recognition rate of LBP-SVM classifier compared to AF-SVM

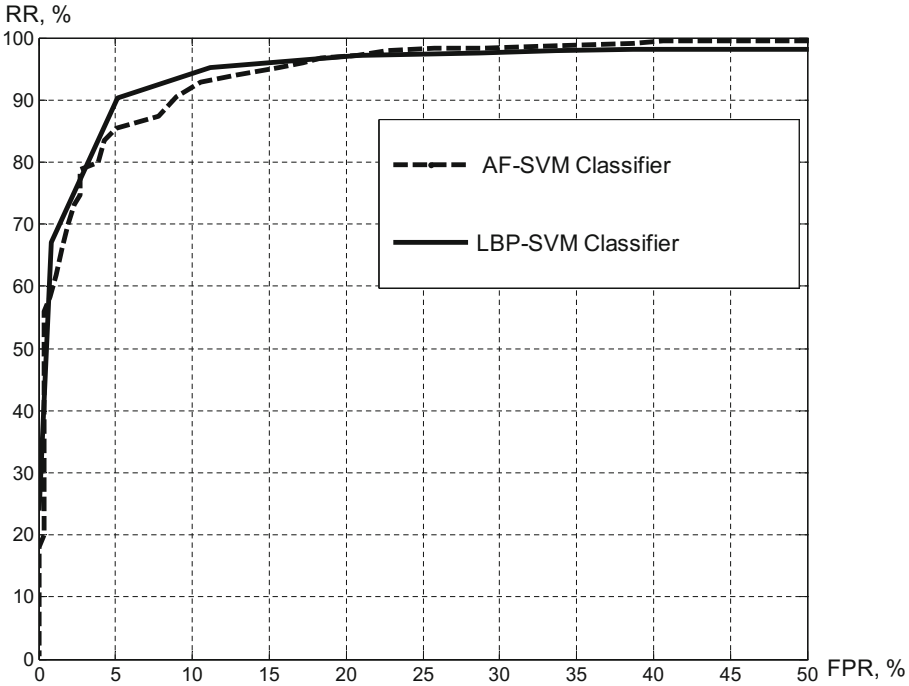| Algorithm / Parameter | AF-SVM | | LBP-SVM | |
|---|---|---|---|---|
| Recognition rate | True | False | True | False |
| Classified as "male", % | 90.6 | 9.4 | 90.2 | 9.8 |
| Classified as "female", % | 91 | 9 | 94.3 | 5.7 |
| Total classification rate, % | **90.8** | 9.2 | **92.3** | 7.7 |

**Fig. 4.** ROC-curves for LBP-SVM and AF-SVM classifiers

Experimental results show that utilization of LBP features for gender recognition improves overall performance by 1.5 % allowing to acquire more than 92 % accuracy.

## 3   Age Estimation

A lot of research in the area of age classification has been done over last few years [27–32]. The proposed age estimation algorithm realizes multiclass classification approach (Fig. 5) where for each age (from 1 to N) a binary classifier is constructed deciding whether a person on input image looks older than the given age or not. 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. Additionally image normalization was performed by histogram equalization procedure. Transformation to LBP feature space and SVM training procedure are used for binary classifier construction. To predict direct age binary classifier outputs are statistically analyzed and the most probable age becomes the algorithm output.

Training and testing require a huge enough color image database. We used state-of-the-art image databases MORPH [33], FG-NET [34] and our own RUS-FD database of real-life test images which low (60 × 60 pixels on each face) resolution (Table 3). Faces on the images were detected automatically by AdaBoost face detection algorithm.
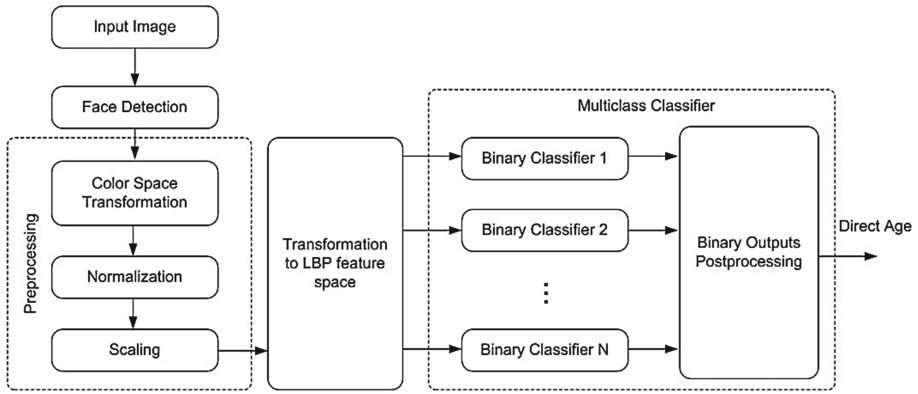
**Fig. 5.** LBP-SVM age estimation algorithm block diagram

**Table 3.** Face databases for age estimation algorithms learning and testing

| Database name | Total number of images using | Range of ages | Age distribution |
|---|---|---|---|
| MORPH | 6513 | 16–75 | Non-uniform |
| FG-NET | 841 | 1–70 | Non-uniform |
| RUS-FD | 6900 | 15–60 | Uniform, 150 faces on each age |

To test age estimation algorithms performance standard metrics were calculated:

- Mean Absolute Error (MAE) – mean absolute difference between estimated and real ages.
- Cumulative Score (CS) – the probability that estimated age lies within an interval $dx$ from real age.
- Probability Density Function of age estimation error.

To estimate the proposed algorithm in real-life situation testing firstly performed on FG-NET database. Age on FG-NET database was marked manually by a group of experts to compare subjective estimation with the algorithm performance. The corresponding dependences for LBP-SVM algorithm simulation are presented in Figs. 6, 7, and 8.

The proposed algorithm shows results comparable to the subjective evaluation in a range of ages from 20 to 35 years. The average absolute error in this range is about 6 years old. Accuracy of LBP-SVM algorithm decreases on senior ages because of MAE grows. In this range (45–60 years), the proposed algorithm yields an expert evaluation approximately 10–15 years in terms of average error.

Cumulative score shows that around 40 % of estimations have less than 5 years deviation from true age and 70 % - less than 10 years deviation. Subjective evaluation curve in Fig. 7 give us the possible limit for future age estimation algorithm improvement.

Analysis of the error probability density function shows that the proposed algorithm has close to symmetric error distribution. Objective results are not inclined to over-estimate the true age, which is typical for the evaluation of experts.
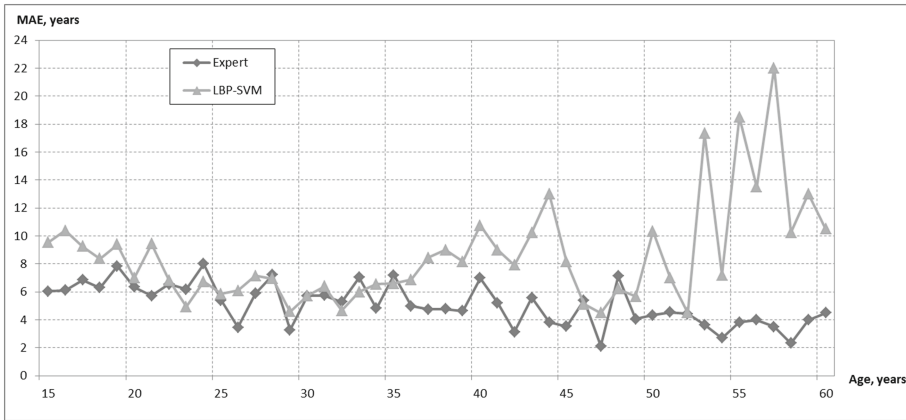


**Fig. 6.** MAE on FG-NET database for the proposed age estimation classifier
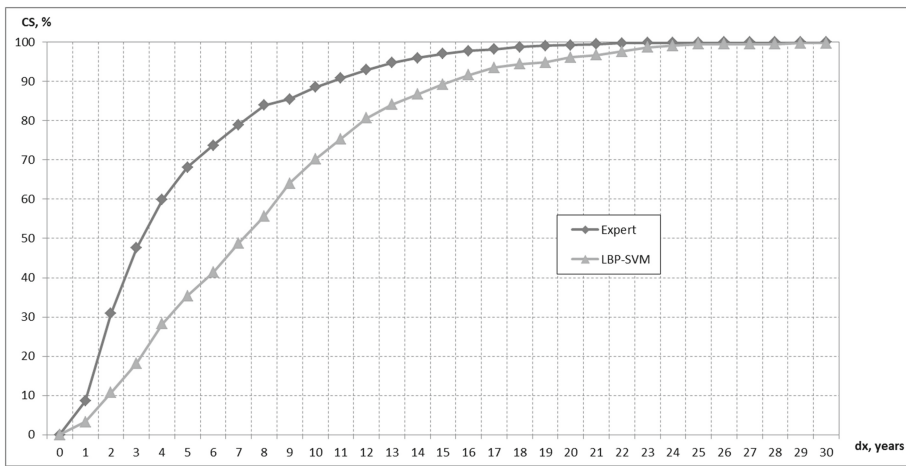


**Fig. 7.** CS on FG-NET database for the proposed age estimation classifier

MAE and CS comparison for LBP-SVM algorithm on different test databases is presented in Fig. 9 and Fig. 10.

Total MAE score of LBP-SVM algorithm on RUS-FD database is 6.94, MORTH database – 7.29, FG-NET database – 7.47. Subjective estimation MAE is 4.2 indicating that the proposed algorithm still needs much improvement to show results comparable to a human. The possible ways to improve the accuracy of age classifier are feature set
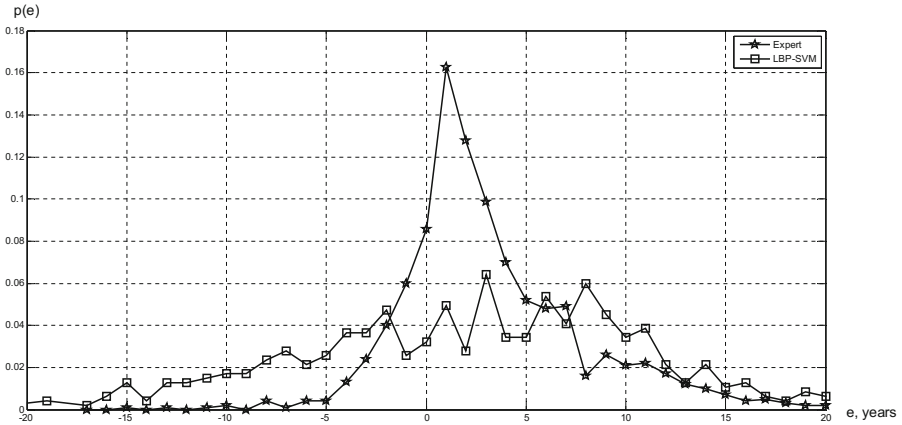
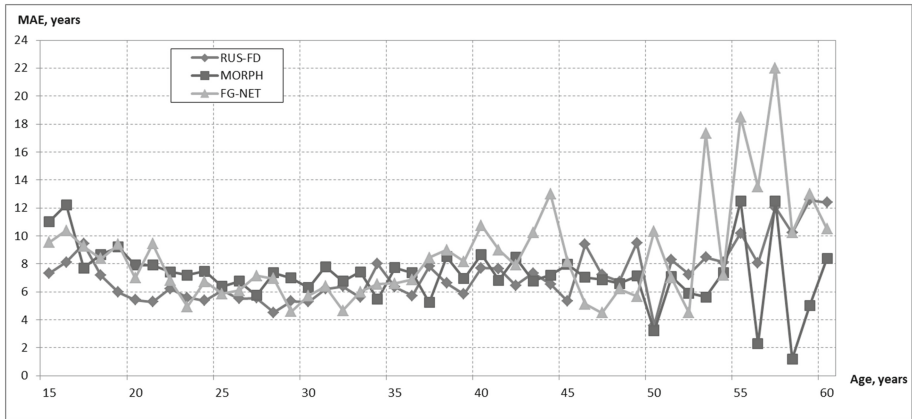**Fig. 8.** Error probability density function on FG-NET database



**Fig. 9.** MAE comparison on different databases for LBP-SVM algorithm

expansion (utilization of a combination of different feature transforms), cost-sensitive SVM learning procedure utilization, pre-processing and post-processing steps efficiency improvement.

## 4   Overall Performance Comparison

The proposed audience analysis system is compared to its commercial analog – Intel Audience Impression Metrics Suite (Intel AIM Suite). Experimental setup was the following: an input video stream from IP-camera (Axis M1014) was split into two and analyzed simultaneously by Intel AIM Suite and by the proposed system.

During the experiment a group of people including men and women have been walking in front of the camera imitating difficult situations of movement such as partial
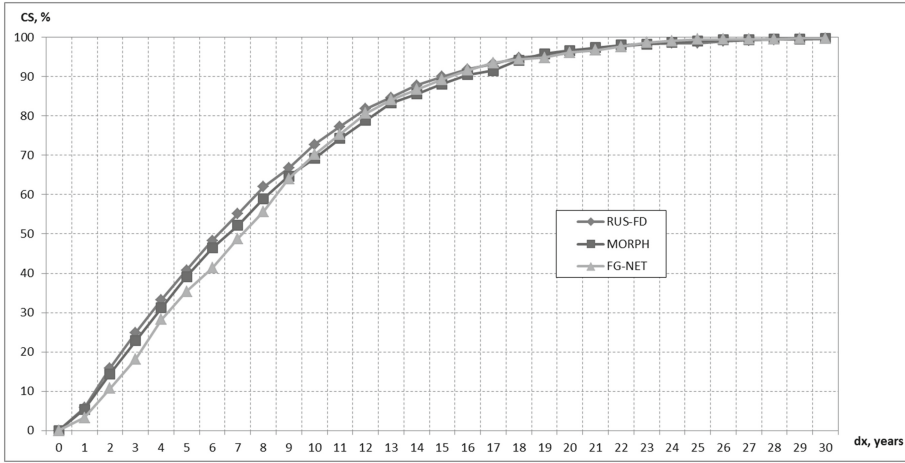
**Fig. 10.** CS comparison on different databases for LBP-SVM algorithm

occlusion and temporary disappearance. The following metric was proposed to compare algorithms performance (Eq. 3):

$$K\frac{D}{N},\tag{3}$$

where $D$ is the total number of misclassified objects on testing video sequence, and $N$ – the total number of frames. Testing results are presented in Table 4. Experimental results show that Intel AIM Suite seriously overestimates the number of people during people count while the proposed system has higher classification accuracy.

**Table 4.** Audience analysis system comparison results

| System / Parameter | Proposed | Intel AIM Suite | Ground Truth |
|---|---|---|---|
| $K$ | 0.29 | 0.6 | - |
| People count | 16 | 72 | 11 |
| Men count | 15 | 63 | 9 |
| Women count | 1 | 9 | 2 |

## 5 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 92 % accuracy.

The noted features allow applying the proposed system in various spheres of life: places of mass stay of people (stadiums, theaters and shopping centers), transport knots (airports, railway and auto stations), digital signage network optimization, etc.

# References

1. Alpaydin, E.: Introduction to Machine Learning. The MIT Press, Cambridge (2010)
2. Sammut, C., Webb, G.I.: Encyclopedia of Machine Learning. Springer, New York (2011)
3. Li, S.Z., Anil, K.J.: Handbook of Face Recognition. Springer, London (2005)
4. Kriegman, D., Yang, M.H., Ahuja, N.: Detecting faces in images: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **24**(1), 34–58 (2002)
5. Hjelmas, E.: Face detection: a survey. Comput. Vis. Image Underst. **83**(3), 236–274 (2001)
6. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: a literature survey. ACM Comput. Surv. (CSUR) **35**(4), 399–458 (2003)
7. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer, London (2010)
8. Makinen, E., Raisamo, R.: An experimental comparison of gender classification methods. Pattern Recogn. Lett. **29**(10), 1544–1556 (2008)
9. Tamura, S., Kawai, H., Mitsumoto, H.: Male/female identification from 8 to 6 very low resolution face images by neural network. Pattern Recogn. Lett. **29**(2), 331–335 (1996)
10. Khryashchev, V., Priorov, A., Shmaglit, A.L., Golubev, M.: Gender recognition via face area analysis. In: Proceedings of the World Congress on Engineering and Computer Science, Berkeley, USA, pp. 645–649 (2012)
11. Fu, Y., Huang, T.S.: Age synthesis and estimation via faces: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **32**(11), 1955–1976 (2010)
12. Sung, K.K., Poggio, T.: Example-based learning for view-based human face detection. IEEE Trans. Pattern Anal. Mach. Intell. **20**, 39–51 (1998)
13. Maydt, J., Lienhart, R.: Face detection with support vector machines and a very large set of linear features. In: IEEE ICME 2002, Lousanne, Switzerland (2002)
14. Roth, D., Yang, M.-H., Ahuja, N.: A SNoW-based face detector. In: Solla, S.A., Leen, T.K., Müller, K.-R. (eds.) Advances in Neural Information Processing Systems 12 (NIPS 12), pp. 855–861. MIT Press, Cambridge (2000)
15. Juell, P., Marsh, R.: A hierarchical neural network for human face detection. Pattern Recogn. **29**, 781–787 (1996)
16. Rowley, H.A., Baluja, S., Kanade, T.: Neural network-based face detection. IEEE Trans. Pattern Anal. Mach. Intell. **20**, 23–38 (1998)
17. Lin, S.H., Kung, S.Y., Lin, L.J.: Face recognition/detection by probabilistic decision-based neural network. IEEE Trans. Neural Netw. **8**, 114–132 (1997)
18. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of International Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 511–518 (2001)
19. Yilmaz, A., Javed, O., Shah, M.: Object tracking: a survey. ACM Comput. Surv. **38**, 13 (2006)
20. Comaniciu, D., Ramesh, V., Andmeer, P.: Kernel-based object tracking. IEEE Trans. Patt. Analy. Mach. Intell. **25**, 564–575 (2003)
21. Shi, J., Tomasi, C.: Good features to track. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 593–600 (1994)
22. Tao, H., Sawhney, H., Kumar, R.: Object tracking with bayesian estimation of dynamic layer representations. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 75–89 (2002)

23. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of Imaging Understanding Workshop, pp. 121–130 (1981)
24. Da, B., Sang, N.: Local binary pattern based face recognition by estimation of facial distinctive information distribution. Opt. Eng. **48**(11), 117203-1–117203-7 (2009)
25. Phillips, P.J.: The FERET evaluation methodology for face recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10), 1090–1104 (2000)
26. Burges, C.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. **2**, 121–167 (1998)
27. Sung, E.C., Youn, J.L., Sung, J.L., Kang, R.P., Jaihie, K.: A comparative study of local feature extraction for age estimation. In: IEEE International Conference on Control Automation Robotics & Vision (ICARCV), pp. 1280–1284 (2010)
28. Thukral, P., Mitra, K., Chellappa, R.: A hierarchical approach for human age estimation. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1529–1532 (2012)
29. Guodong, G., Guowang M.: Human age estimation: What is the influence across race and gender. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 71–78 (2010)
30. Zhen, L., Yun, F., Huang, T.S.: A robust framework for multiview age estimation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 9–16 (2010)
31. Guodong, G., Xiaolong, W.: A study on human age estimation under facial expression changes. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2547–2553 (2012)
32. Hee, L.W., Jian-Gang, W., Wei-Yun, Y., Xing, L.C., Yap, P.T.: Effects of facial alignment for age estimation. In: IEEE International Conference on Control Automation Robotics & Vision (ICARCV), pp. 644–647 (2010)
33. Ricanek, K., Tesafaye, T.: MORPH: a longitudinal image database of normal adult age-progression. In: IEEE 7th International Conference on Automatic Face and Gesture Recognition, pp. 341–345 (2006)
34. The FG-NET Aging Database. http://www.fgnet.rsunit.com/, http://wwwprima.inrialpes.fr/FGnet/