

Real-time high-resolution detection approach considering eyes and its states in video frames through intelligence-based representation

S. N. Yahyavi¹ · A. H. Mazinan² · M. Khademi³

Received: 25 March 2015 / Accepted: 29 April 2016 / Published online: 12 May 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract The research relies on the efficient real-time high-resolution detection approach through intelligence-based representation to consider the human eyes in connection with its states that can be a distinction between the open and closed ones. In a word, the subject behind the research is achieved using the principal component analysis in association with the artificial neural networks. First, the face needs to be detected using the AdaBoost technique. Then, after a number of processing steps, the areas of the mouth, the left and right nostrils as well as the eyes should be identified. Moreover, the eyes are separated from the face and passed to the multi-layer perceptron for the purpose of classifying the results. Having identified the features of the open and closed eyes through the aforementioned principal component analysis in line with the present intelligence-based technique, the outcome is realized to identify the states of the eyes from frames taken by webcams. The proposed approach can be used to indicate the left and right click on a computer by opening and closing left and right eyes, as well. A number of frames from various sources in the different lighting con-

ditions are carried out to demonstrate the robustness of the approach in coping with any situation.

Keywords High-resolution detection approach · Human eyes and its states · Intelligence-based representation · AdaBoost technique · Principal component analysis

Introduction

The face and its components detection, in particular, the eyes are known as one of the challenging issue in the area of machine vision in the last two decades. There are various researchers that are working in this way to propose potential solutions. The detection of the eyes states has many practical uses including security and safety systems, control systems, driver behavior and human computer interaction (HCI). It should be noted that a number of techniques are presented in the area of the face and eye detection as well as its tracking. These methods are realized based on the different features such as composition, shape and colour or a combination of them. One of the methods for the face detection is to use the skin colour, which is efficient in widespread applications up to now [1–6]. Another method for the face detection is to use facial features extracted by the principal component analysis. This is widely used in the face detection and the corresponding recognition systems [7]. There is also motion analysis to find the head and the face in video frames [8]. It is also possible to take advantage of recursive nonparametric discriminant analysis (RNDA) to extract features of the eyes and pass it to a classifier for detection [9].

Other techniques have been also suggested to detect and classify the states of eyes including distinctive features of eyes, such as corners and coloured sections of the iris. And

✉ A. H. Mazinan
mazinan@azad.ac.ir; ahmazinan@gmail.com;
ah_mazinan@yahoo.com

¹ Department of Electronics Engineering, South Tehran Branch, Islamic Azad University (IAU), No. 209, North Iranshahr St, P.O. Box 11365/4435, Tehran, Iran

² Department of Control Engineering, Faculty of Electrical Engineering, South Tehran Branch, Islamic Azad University (IAU), No. 209, North Iranshahr St, P.O. Box 11365/4435, Tehran, Iran

³ Department of Applied Mathematics, Faculty of Technical and Engineering, South Tehran Branch, Islamic Azad University (IAU), No. 209, North Iranshahr St, P.O. Box 11365/4435, Tehran, Iran

also variance filter analysis can be realized for eye detection and tracking utilising vertical projection for finding the exact location. The distinctive features of eyes can be realized to detect the eyes and its states, while corners of the eyes and edges of the iris in line with its center should be identified to locate the upper eyelid. The distance between the iris and upper eyelid is important to establish if the eye is open or closed and how far it is closed. A method based on line to line analysis of the face area is considered for eye blink detection [10, 11]. For this, the face area is scanned and analysed line by line from top to bottom to locate the position of eyes. After detecting the eyes, its profile is matched against a database of known states of the eyes to determine these ones. Light flow is used to analyse future frames for detection of eyelid motion. A very fast and reliable method for face detection in colour pictures is considered through the AdaBoost method. In this method, three original ideas are realized, while the first one is a representation of pictures as integral pictures so that the required features can so quickly be calculated [12–14].

Instead of relying on intensity of light, it uses the Haar-like features. Using integral pictures, the speed of processing is significantly increased by reducing the number of calculations. Each Haar-like feature can be calculated at any scale in any location in constant time. The key point is to reduce the number of features using the AdaBoost. In any picture, the number of Harr-like features is much bigger than the number of pixels. By focusing on a small number of critical features, the performance can largely be

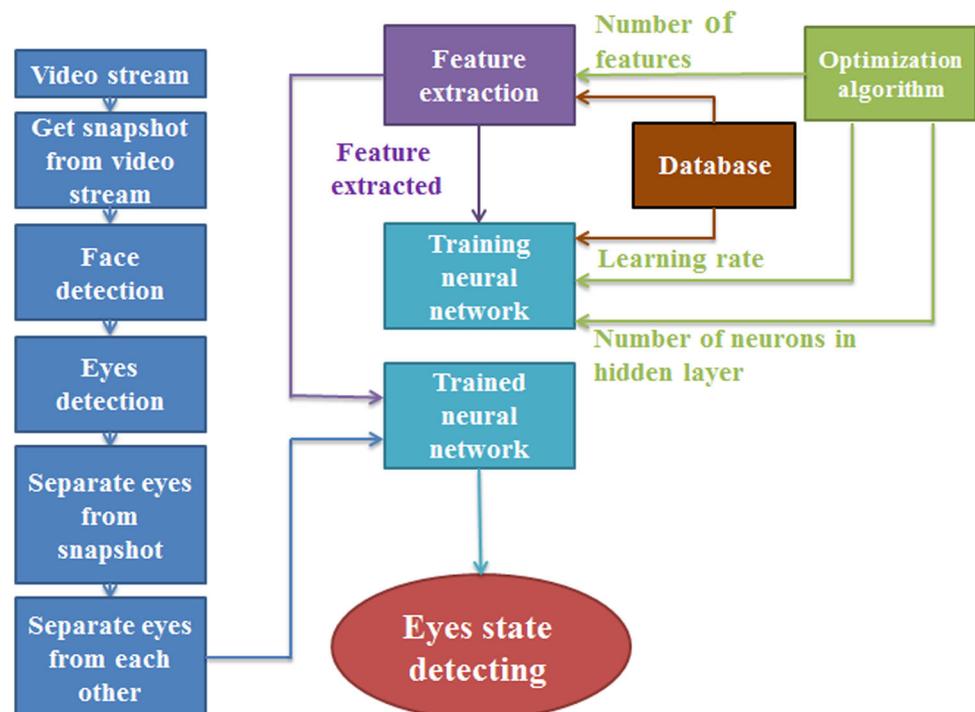
improved. The feature selection can be improved by modifying the AdaBoost. It is to note that the weak learner is constrained by reducing the weak classifier to one single feature. The selection of each new weak classifier can be viewed as a feature selection method in the boosting process. This provides an effective generalised algorithm for learning [15–17]. The outcomes are to focus on the areas of the picture, which have important features. By combining more complex classifiers in a cascade structure, the speed of detector is increased and focused on the regions of the picture. This helps us to determine in which areas of the picture objects may occur and focus on the more complex processing for these regions. Using the false-negative rate, almost all objects can be chosen through the filter [18].

The rest of the research is organised as follows: The proposed high-resolution detection approach and its experimental results are presented in “The proposed high-resolution detection approach and its experimental results”. Finally, the research concludes the investigated outcomes in “Conclusion”.

The proposed high-resolution detection approach and its experimental results

The proposed high-resolution detection approach is schematically illustrated in Fig. 1. It is to note that designing a system in capable of automatic and real-time eye detection and its

Fig. 1 The schematic diagram of the high-resolution detection proposed approach



states in the different light conditions is always difficult. The eyes are complicated by two-dimensional pictures with a number of varied states and features depending on emotional state of the subject. This research is to introduce an accurate and useful approach to detect the eyes and its states, carefully. To reduce the number of calculations involved in this complicated analysis, the principle component analysis (PCA) in association with the artificial neural networks is realized.

In fact, the output of the aforementioned PCA is passed to the artificial neural network for purpose of classifying the results to the open and closed eyes.

The database

To compare the effectiveness of different algorithms, a systematic and complete database is first necessary to be evaluated. This is made up of pictures of eyes, which are taken from various angels at different lighting conditions in two states including the open and closed ones. Because there is no readily available database for this purpose, the team created a unique database of pictures to be applicable. The present database consists of 640 pictures taken from the eyes of 160 subjects in both open and closed states. This includes 135 men and also 25 women. Moreover, 27 people out of 160 subjects wore spectacles. They are in all different ages. The pictures are taken in full colour at resolution of 640×480 pixels covering the whole face. Using MATLAB programming language, these are converted to greyscale pictures of eyes only at a resolution of 81×81 pixels. As are obvious in Fig. 2, these are classified into two classes of open and closed eyes.

The feature extraction and the artificial neural network training

To realize the approach proposed here, the feature extraction in connection with the artificial neural network training should be implemented. In one such case, the pictures are split into two groups based on the state of the eyes including the open and closed ones. From each group, randomly, 80 % (256 pictures) are selected to train and 20 % (64 pictures) are also selected to test. Because of this selection technique, there is no possibility of the same picture being used for training and testing. Using the PCA, the eigenvectors and eigenvalues of the present pictures ate to be extracted. From the point of these values, the PCA is asked to select 260 features to pass to the artificial neural network. A multi-layer perceptron (MLP) neural network is employed to carry out the analysis. Hence, the matrix of input regarding the MLP for training consists of 512 rows and 260 columns. Moreover, the output regarding

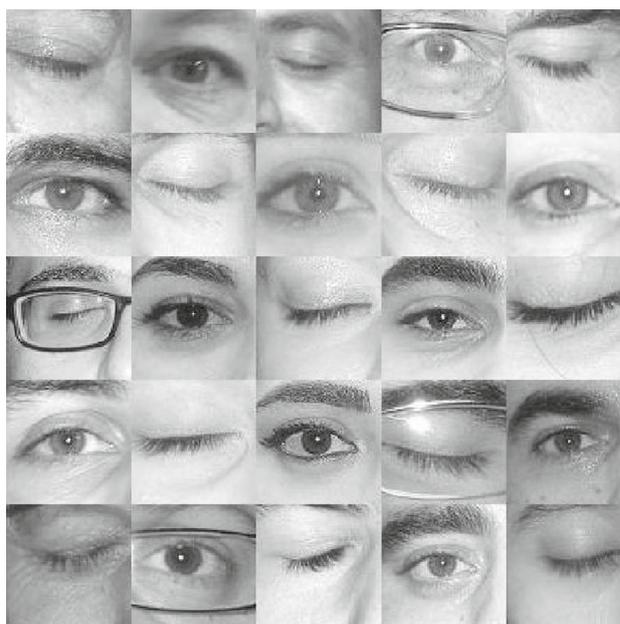


Fig. 2 The samples of database

Table 1 The parameters of the artificial neural network

EPOCH	Learning rate	Hidden layer	Data for train (%)	Sample
500	0.05	20	90	640

the MLP consists of two neurons, which determine states of open and closed eyes through a binary code.

To determine the optimal hidden layer and learning rate in the MLP, the Imperialist competitive algorithm (ICA), the particle swarm optimisation (PSO) and the genetic algorithm (GA) are realized. The data have been normalised by sigmoid function to an output of between 0 and 1. The MLP learns the appropriate bias and weights from the training sample of 512 pictures. The neural network parameters, which used for the simulation, are tabulated in Table 1. One important point to note at this stage is that with each epoch the mean square error rate reduces as illustrated in Fig. 3, where the true-positive rate in line with the false-positive rate in the form of the region of convergence (ROC) is shown in Fig. 4. As expected to reach, while the pictures that are used in the process of training are given as the input, the outcomes for each one of the classes are 100 % accurate. In the process of testing, the input matrix has 128 rows and 260 columns, respectively. The output for each row is a figure between 0 and 1 as detailed above.

The statistics for the tests are tabulated in Table 2. It should be noted that the class is taken as the state of the eye including 1 being closed and 2 being open. The sensitivity-true-positive

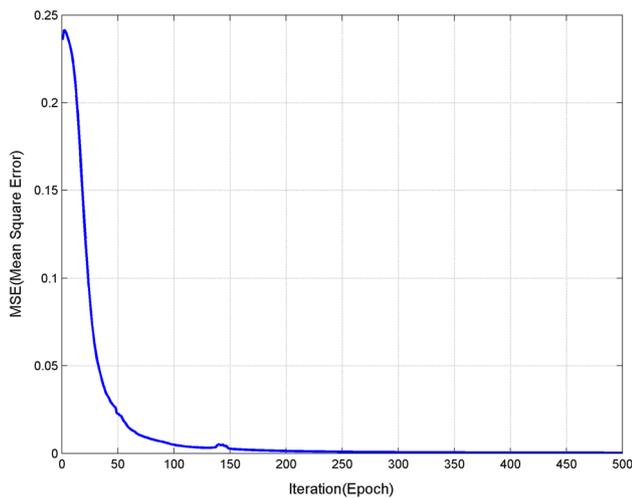


Fig. 3 The outcome of the MSE

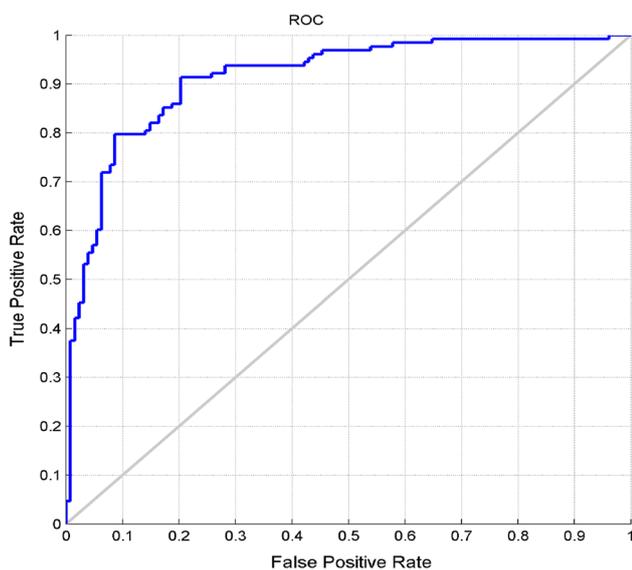


Fig. 4 The outcome of the ROC

rate (SEN – TPR), the specificity-true-negative rate (SPC – TNR), the accuracy (ACC) and the positive predictive value (PPV) as the performance are taken as:

$$\text{SEN} - \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{P} \quad (1)$$

Table 2 The confusion matrix for each class

Class	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy	Performance
1	61	55	3	9	87.14	94.82	90.62	95.31
2	55	61	9	3	94.82	87.14	90.62	85.93
Average					90.98	90.98	90.62	90.62

$$\text{SPC} - \text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} = 1 - \text{FPR} \quad (2)$$

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

where TP is taken as the number of the true-positive predictions, TN is taken as the number of the true-negative predictions, FP is taken as the number of the false-positive predictions and, finally, FN is taken as the number of the false-negative predictions.

The face detection

The face detection is important to be used in the process of the proposed high-resolution detection approach. To reduce the effect of different lighting conditions, the normal red–green–blue (RGB) colour model should be replaced by the normalised red–green (RG) colour model. This transformation idea is realized by the following:

$$r = R / (R + G + B) \quad (5)$$

$$g = G / (R + G + B) \quad (6)$$

In this method, the geometric properties of the face are used to determine five points, where they are taken for the two eyes, the two nostrils and the mouth, respectively. A sample of pictures used to be evaluated through the proposed approach is illustrated in Fig. 5, where eye regions are extracted to be passed to the MLP for processing as shown in Fig. 6.

The investigated outcomes

Four people are first chosen and a video of them in various states with one or two eyes in either open or closed states has been taken to cover all the possible positions to be evaluated. The outcomes of the sample picture from the proposed approach are illustrated in Fig. 7. In these pictures, the system has identified the eyes regions by drawing a frame around each one of the eyes. It has also identified the states of the eyes by colour coding including the red for closed and the green for open ones.



Fig. 5 The samples of the pictures

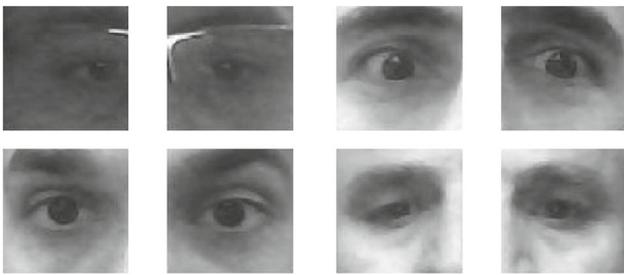


Fig. 6 The samples of the eyes extraction

The proposed approach for the detection of eyes and its states are carried out by 25 new subjects using continuous video frames. Each video has 1000 frames. Each subject is dealt with to blink 15 times during the process. The light source is placed above the head of subjects. The acquired results of four subjects, which are randomly selected, are tabulated in Table 3.

The results indicate that the proposed approach is obviously carried out with the high performance. The investigated results are considered to be compared with two benchmarks including the Sirohey and the Lalonde, as well. The results are tabulated in Table 4, where the outcomes confirm that the proposed one is favourably behaved with respect to the benchmarks.

Conclusion

A real-time high-resolution detection approach is considered in the present research to deal with the human eyes in connection with its states through intelligence-based representation. The proposed approach is able to address a distinction between the open and closed states. There are a number of processing steps to be carried out including

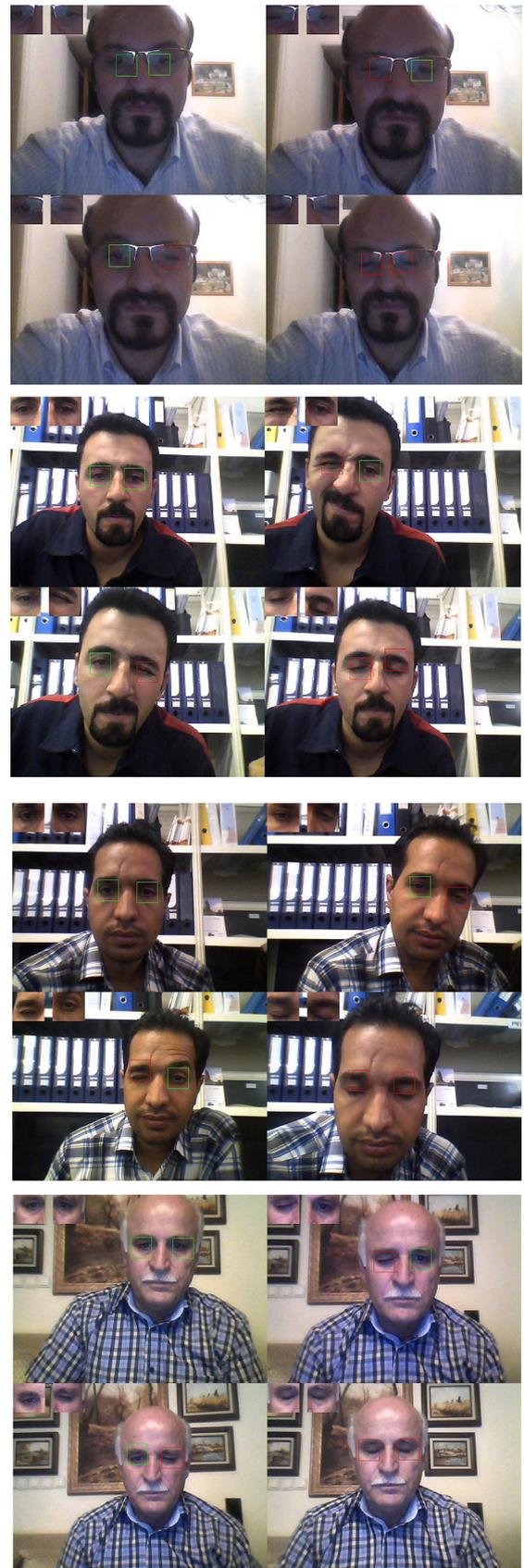


Fig. 7 The results of the detection

Table 3 The results of the proposed algorithm

Video stream	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	
Number of blinking	15	15	15	15	15	15	15	15	15	15	
Number of right eye blinking	5	5	5	5	5	5	5	5	5	5	
Number of left eye blinking	5	5	5	5	5	5	5	5	5	5	
Number of both eyes closed	5	5	5	5	5	5	5	5	5	5	
Number of total blinking detected	14	15	16	13	15	15	14	15	15	16	
Number of right blinking detected	5	5	5	5	5	5	4	5	4	5	
Number of left blinking detected	4	5	5	4	4	4	3	5	5	4	
Number of both eyes blinking detected	5	5	6	4	6	5	4	5	5	4	
Correct right eye blinking detected (%)	100	100	100	100	100	100	80	100	80	100	
Correct left eye blinking detected (%)	80	100	100	80	80	80	60	100	100	80	
Correct both eyes closed (%)	100	100	80	80	80	100	80	100	100	80	
Performance (%)	93.33	100	93.33	86.66	86.66	93.33	73.33	100	93.33	86.66	
Video stream	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	All of samples
Number of blinking	15	15	15	15	15	15	15	15	15	15	1200
Number of right eye blinking	5	5	5	5	5	5	5	5	5	5	400
Number of left eye blinking	5	5	5	5	5	5	5	5	5	5	400
Number of both eyes closed	5	5	5	5	5	5	5	5	5	5	400
Number of total blinking detected	15	15	14	15	15	15	15	14	14	15	1093
Number of right blinking detected	5	5	5	5	5	4	5	4	5	5	383
Number of left blinking detected	4	5	5	5	4	4	3	4	5	5	367
Number of both eyes blinking detected	5	4	4	5	5	5	5	4	5	4	343
Correct right eye blinking detected (%)	100	100	100	100	100	80	100	80	100	100	95.75
Correct left eye blinking detected (%)	80	100	100	100	80	80	60	80	100	100	91.75
Correct both eyes closed (%)	100	80	80	100	100	100	100	80	100	80	85.75
Performance (%)	93.33	93.33	93.33	100	93.33	86.66	86.66	80	100	93.33	91.08

Table 4 The comparison of the proposed approach with the two benchmarks

The approach's titles	The closed eye detected	The eye detection in rotating head state	The closed eye tracking	The number of frames processed per second
Sirohey approach [10]	No	Yes	No	Less than 10
Lalonde approach [11]	Yes	Limited (less than 30°)	Yes	25
The proposed approach	Yes	Yes	Yes	12

detection of the face through AdaBoost technique and also identifications of the areas of the mouth, the left and right nostrils as well as the eyes that are all important to be evaluated. It is to note that the features of the open and closed eyes are considered to identify the states of the eyes through the principal component analysis in association with the artificial neural network. Regarding the further works, it is possible to develop the approach investigated here by linking the colour frames around the left and right eyes through the corresponding left and right mouse clicks to deal with a various applications. It is also possible to track the position of the iris and link it to the movement of pointer on the

screen. Both these ideas can improve the human–computer interaction.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- González-Ortega D, Díaz-Pernas FJ, Antón-Rodríguez M, Martínez-Zarzuela M, Díez-Higuera JF (2013) Real-time vision-based eye state detection for driver alertness monitoring. *Pattern Anal Appl* 16(3):285–306
- Espinosa J, Roig AB, Pérez J, Mas D (2015) A high-resolution binocular video-oculography system: assessment of pupillary light reflex and detection of an early incomplete blink and an upward eye movement. *BioMedical Engineering*, December 2015, pp 14–22
- Kuehlkamp A, Franco CR, Comunello E (2014) An evaluation of iris detection methods for real-time video processing with low-cost equipment. *Information Sciences and Systems*, pp 105–113
- Chopra P, Yadav SK (2016) Fault detection and classification by unsupervised feature extraction and dim. *Complex Intell Syst* (in press)
- Fathi A, Manzuri MT (2004) Eye detection and tracking in video streams. *IEEE International Symposium on Communications and Information Technology*
- Kovac J, Peer P, Solina F (2003) Human skin color clustering for face detection. *IEEE EUROCON Computer as a Tool*
- Pentland A, Moghaddam B, Starner T (1994) View-based and modular eigenspaces for face recognition. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*
- Magee JJ et al (2004) EyeKeys: a real-time vision interface based on gaze detection from a low-grade video camera. *CVPRW Conference on Computer Vision and Pattern Recognition Workshop*
- Wang P, Ji Q (2007) Multi-view face and eye detection using discriminant features. *Comput Vis Image Underst* 105(2):99–111
- Sirohey S, Rosenfeld A, Duric Z (2002) A method of detecting and tracking irises and eyelids in video. *Pattern Recogn* 35(6):1389–1401
- Lalonde M et al (2007) Real-time eye blink detection with GPU-based SIFT tracking. *Fourth Canadian Conference on Computer and Robot Vision*
- Zhou Z-H, Geng X (2004) Projection functions for eye detection. *Pattern Recogn* 37(5):1049–1056
- Viola P, Jones MJ (2004) Robust real-time face detection. *Int J Comput Vis* 57(2):137–154
- Papageorgiou CP, Oren M, Poggio T (1998) A general framework for object detection. *Sixth International Conference on Computer Vision*
- Kinh T, Viola P (2000) Boosting image retrieval. *IEEE Conference on Computer Vision and Pattern Recognition*
- Osuna E, Freund R, Girosi F (1997) Training support vector machines: an application to face detection. *IEEE Conference on Computer Vision and Pattern Recognition*
- Schapiro RE, Freund Y, Barlett P, Lee WS (1997) Boosting the margin: a new explanation for the effectiveness of voting methods. In: *Proceedings of the fourteenth international conference on machine learning (ICML)*, Nashville, Tennessee, USA, 8–12 July 1997
- Aldrian P, Meier U, Pura A (2009) Extract feature points from faces to track eye's movement. *University of Leoben, Austria*