An Efficient face detection based on improved Viola & Jones

Kawther.Benhallou[1], Mokhtar Kech, Abdelaziz Ouamri, Khadidja Benhallou
Signals and Images Laboratory, Department of Electronics,
Faculty of Electrical Engineering, University of Science and Technology ORAN – Mohamed Boudiaf – USTO-MB, B.P. 1505
El Mnaouer, Oran, 31000, Algeria
[1] benhallou-kawther@hotmail.com

***Abstract*— Human face detection plays an important role in many applications such as computer-human interaction, face recognition and video surveillance and so on. In human face detection application a false detection rate and slow detection speed remain major problems. In this paper, we present a method to avoid these problems, by combining a Fast Saliency-based visual attention (FSBVA), skin color filtering (SCF), and Viola and Jones Face detector (VJFD) : After determining all possible face candidate regions using FSBVA with SC, the VJFD is applied to detect Human face and then a discrimination based SCF is used to reject false positive. Our method takes advantage from the visual attention mechanism to rapidly find the attractive face regions and to overcome the problems of false negatives when both VJFD and SC are applied directly in video frames with complex backgrounds. Experimental results show the effectiveness of the proposed method.**

***Index Term*— Face detection, saliency based visual attention, Viola and Jones face detector, skin color filtering.**

I. Introduction

Human face detection plays an important role in many applications such as computer-human interaction, face recognition and video surveillance and so on. The problem of face detection refers to determining whether or not there are any faces in a given image [01]. There have been various approaches proposed for face detection, which could be generally classified into four categories [02]: template matching based methods, feature-based methods, knowledge-based methods, and learning based methods. Template matching based method means the final decision comes from the similarity measurement between input image and the template. It is scale-dependent, rotation-dependent and computational expensive. Feature-based methods use low-level features such as intensity [3], color [4], edge, shape [5] , and texture to locate facial features, and further, find out the face location. Knowledge-based methods [6] detected an isosceles triangle (for frontal view) or a right triangle (for side view). Learning based methods use a lot of training samples to make the classifier to be capable of judging face from non-face.

Recently, a lot of research is being done in the vision community to accurate face detector in real work application, in particular, the seminal work by viola and Jones [7]. The Viola and Jones face detector has become the defacto standard to built successful face detection in real time, however, it produces a **high false positive** (detecting a face when there is none) and **false negative rate** (not detecting a face that's present) when directly applied to the input image. To deal with this problem, a various improvements have been proposed, such as using skin color filters (whether pre- filtering or post-filtering) to provide complementary information in color images. Though many experimental results have demonstrated the feasibility of combining **SCF** with **VJFD to reduce false positive, both methods suffer** from **high false negative rate** as some face regions may be ignored by detector, **when directly applied to the input image with complexes background**[10][13][17]. This can significantly **decrease the accuracy of any face application domains,** attributed to the fact that, when a face region **is missed**, the following stages of the system **cannot retrieve** the missed face. Therefore, false negative determine the eventual success or failure of the subsequent stages.

In this paper, we propose a way to overcome this limitation by exploring the **advantages of the fast saliency-based visual attention** to **select the image regions** which may contain **the most relevant parts** of video frames **i.e., rapidly finding the attractive face regions**. The proposed method uses **fast spatiotemporal saliency**, combined with skin color pre-filtering to select candidate regions, and then applies **Viola and Jones face detector VJFD** with **skin color-post-filtering.**

The remainder of this paper is organized as follows, Sect.II, provides an overview of related work. The proposed method is presented in Sect.III. Sect.IV is devoted to the experimental results. Finally, conclusions and perspectives are drawn in Sect.V.

II. RELATED WORK

The Viola and Jones face detector is the first ever face detection framework to provide successful face detection in real time. It contains three main ideas that make it possible to run in real time: the integral image, classifier learning with Adaboost, and the intentional cascade structure [8]. However, it produces a high false positive rate and false negative when directly applied to the input image. Various research contributions have been made to overcome these problems, such as using pre-filtering or post-filtering methods based skin color filter to provide complementary information in color images. In [9] [10] the authors proposed an interesting method to reduce the false detection by using a skin color as a pre- filtering stage prior to the application VJFD. In [11] the authors proposed a hybrid method to reduce false positive in the VJFD by using skin color face post-filtering method in HSV color space. To reduce the effects of lighting, the authors in [12] applied an illumination compensation algorithm in the first step and then, they combine VJFD and the skin color detector to detect face. In [13] the authors proposed a method to reduce the false positive rate and keeping the high detection rate of the VJFD in real applications, by using post-filtering methods based on color-based skin filtering in RGB color space. Though many experimental results have demonstrated the feasibility of combining SCF with VJFD to reduce false detection, both methods suffer from high false negative rate when directly applied to the input image. This can significantly decrease the accuracy of any face application domains, attributed to the fact that, when a face region is missed, the following stages of the system cannot retrieve the missed face. Our goal is to propose an efficient face detection method which guarantees reducing false negative and false positive rate on the work of Viola and Jones face. The motivation of our method takes advantage of visual attention mechanism to rapidly find the attractive face regions and to overcome the problems of false negatives when both VJFD and SC are applied directly in video frames with complex backgrounds.

Computational models of visual attention have been widely soused in modern computer vision and robotics to predict important areas in the visual field, [31], [22]. The researchers have seen a renewed interest in the use of visual attention **to rapidly towards on a small object of interest** in our visual environment and they suggest that **Human face** and text **primarily attract attention** independent of tasks [24][25]. The saliency-based visual attention represents one of the main attempts to simulate this visual mechanism on computer [23]. Several approaches have been proposed in the literature to provide a computational foundation to the notion of visual saliency, which can be broadly classified into a)bottom–up [13] [21], b) top–down methods and c) Bayesian or hybrid model. Bottom-up attention methods use local features in a given image to find image locations, which are considerably different from their neighbors [25].

Recently, several hybrid models of visual attention have been proposed by various researchers. These models attempt to formulate human attention in a Bayesian framework, combining the bottom-up saliency model and the top-down contextual information of object location and appearance [20-26,27]. Usually, Such models, implicitly or explicitly, estimate a probability density of filter responses obtained from local image features (bottom-up saliency) for a given image and combine it with the probability density of object shape (shape prior) and object location (location prior), learned from the training samples in a Bayesian framework. In [31] a bottom-up model of visual attention based on three image features: colors, intensity and orientation in different scales, is proposed. The center-surround operations are implemented by difference of filter responses between two scales to obtain a set of feature maps for a given feature. The feature maps are then normalized and linearly combined to construct the final saliency map. In [20], the authors proposed a bottom-up model based on the use of dissimilarity metric. They defined a graph on the image and employed random walks to compute visual saliency. The saliency measure at a spot is proportional to the frequency of visits at the equilibrium of the random walk. In [18], a hybrid model of visual attention, which attempts to calculate the measure that exploits statistics of natural images, is proposed. The difference of Gaussian filters (DoG) is used for the calculation of local saliency. The image intensity is used as input and the filter responses are estimated by a multivariate generalized Gaussian distribution. However, these models of visual saliency are typically too complex and need a very high computational capacity [22]. It is necessary that these models produce fast responses for high-level applications such as robot vision and video content analysis. To provide a real-time model of visual attention, Nich et al. [22] proposed a simplified version of the model proposed by Zhang et al. [18]. They empirically evaluated the saliency model in saccades control of a camera in social robotics situations.

III. DESCRIPTION OF PROPOSED METHOD

We propose an efficient face detection method which reduces the false negative and false positive rate of the Viola and Jones face detection method. Our method takes advantage of saliency-based visual attention to overcome the problem of false negatives by rapidly finding the attractive face regions in video frames. This method consists of three stages: (1) Pre-selecting candidate face areas, (2) face detection using Viola and Jones, (3) validation with skin color post-filtering. The flowchart in Figure 1 shows the steps followed to decrease the false detection rate and false negative.

A. Pre-selecting candidate face areas

Skin has a characteristic color which is shown to be an effective cue for face detection; however, skin color filtering suffers from false negatives, i.e. miss-detection, when directly applied to the input images. To overcome this problem we propose to use the fast spatiotemporal saliency technique, to rapidly find the attractive face regions in video frames, and combine them with those obtained by skin color filtering. The combination is achieved by merging the binary image masks given by the two methods, using the bitwise OR operator. This first stage is a filter to narrow the search space, reduce the false negatives rate and speed up the VJFD.

Salient object extraction

The attractive salient objects which correspond to the most salient regions of the saliency map are extracted by using the model proposed by Nich et al. [22], then the binary mask image is generated by using simple thresholding to segment the saliency map into salient regions and the background.

In [22] the authors proposed a very fast model of visual attention for robot vision. This model is an approximation to the Bayesian model presented in [18]. The approximation is achieved in the following way:

- Instead of the Difference-of-Gaussian spatial filters, the difference of DoB is used. The DoB are types of box Haar-style filters, a computationally efficient class of filters that have been used with much success recently in visual object classification in the work of Viola and Jones [07].

- Instead of the Generalized Gaussian probability model, the Laplacian probability model is used; the filter impulse response distribution was modeled as a Laplacian distribution with fixed variance.

The algorithm is designed to find regions of an image plane most likely to be useful in unconstrained conditions, i.e., situations where there are a very large number of potential tasks of interest.

After selecting the regions which may contain the most important information, the saliency map is analyzed for the

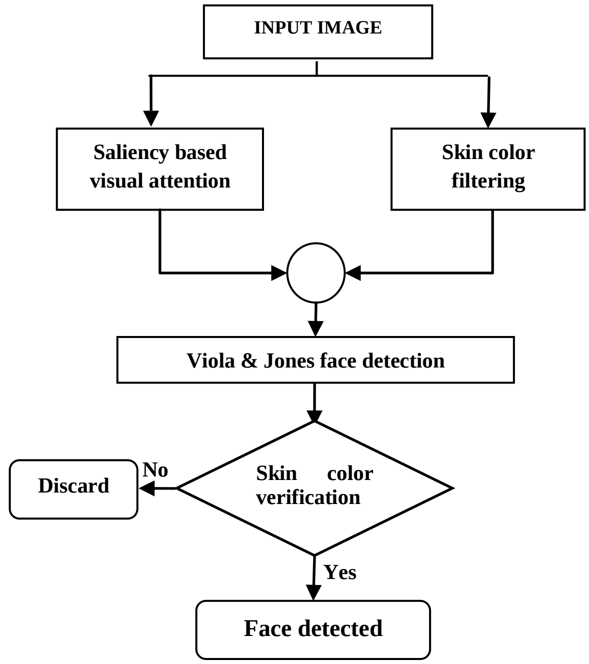


Fig. 1. Flowchart of the proposed method.

segmentation method, inspired by the approach used in [28], converts the saliency map *Is*(x,y) to a binary (black and white) detection and selection of the regions of interest, discarding regions which are too small to contain useful information.

Firstly, the segmentation is used to extract object corresponding to the most salient region of the saliency map. The threshold mask I_B(x,y) according to (1). The threshold value is interactively determined, using the saliency map. Then, morphological operations are used to remove small regions and fill holes.

$$I_B(x) = \begin{cases} 1 & ,I_S(x) \geq threshold \\ 0 & ,I_S(x) < threshold \end{cases} \tag{1}$$

Skin color filtering

Recently the skin color methodologies have been broadly used in a great number of applications as complimentary information to other features. Skin color is an affective cue for face detection since it provides computationally effective, robust information against rotation, scaling and partial occlusions. In general, the final goal of skin color filtering is to build a decision rule, which will discriminate between skin and non-skin pixels. There have been various approaches for skin color filtering summarized in several survey papers [14, 15], and may fall into two main categories [16]: pixel-based skin detection methods and region-based skin detection methods. Pixel-based skin detection methods classify each pixel as skin or non-skin individually, independently from its neighbors. On the other hand, region-based skin detection methods try to take the spatial arrangement of skin pixels into account during the detection stage to enhance the methods performance.

In our proposed method, especially for real-time implementation, the pixel-based skin detection method is chosen for fast processing. We also use a color compensation step [29] prior to skin color filtering, to reduce the effects of lighting. A pixel with color components (R, G, B) is detected as skin if the conditions given in (2) below hold. The second line in (2) ensures that RGB components must not be close together, which ensures greyness elimination. The third line in (2) ensures that R and G components must not be close together, which must be true for fair complexion [14].

$$\begin{gathered} R > 95 \ \& \ G > 40 \ \& \ B > 20 \ \& \\ \max\{R,G,B\} - \min\{R,G,B\} > 15 \ \& \\ |R - G| > 15 \ \& \ R > G \ \& \ R > B. \end{gathered} \tag{2}$$

B. Viola and Jones face detection

The purpose of this section is to classify the areas, into faces or non faces. The Viola and Jones face detector is performed locally on all selected bounding boxes around connected pixel regions on an image. This allows eliminate fake faces with low saliency, and help to decrease the false positives in face detection. Viola & Jones [7] have presented a face detection method based on an over-complete set of Haar-like features (fig.3) which are calculated in scaled analysis windows. The rectangular Haar-like features are sensitive to edges, bars and other similar structures in the image and they are computed using an efficient method based on the integral image concept.

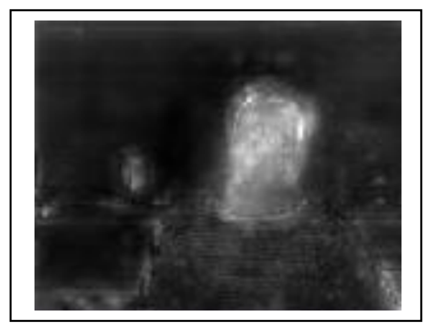
After calculation of a huge number of features for each analysis window, the AdaBoost algorithm is used for combining a small number of these features to form an effective classifier.

C. Skin color verification

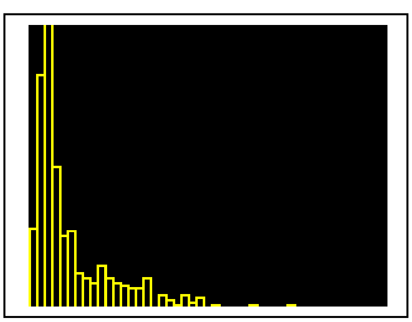
As mentioned, Viola and Jones face detector has some substantial limitations, as a high false positives rate. In order to solve problem, Skin color post- filtering is used to verify if there are sufficient number of skin pixels in a face candidates sub-windows detected by VJFD; all sub-windows with a size smaller than Ω = 300 pixels are discarding. Verification with skin color filtering is illustrated in Figure.5.

Fig. 2. Detection of most salient pixels in video frames.

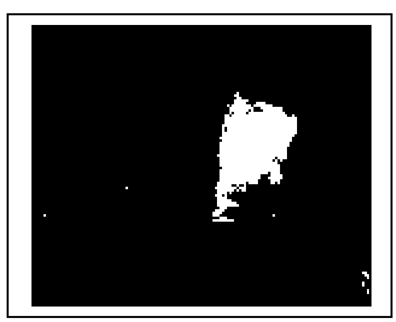
(a) The original image



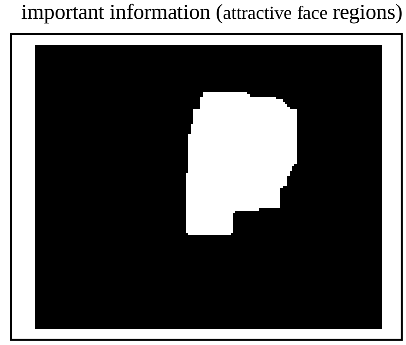
(b) The result of extracting the most important information (attractive face regions)



(c) The histogram of the block-wise saliency map



(d) thresholding the saliency shows the most salient pixel on the image



(e) image after applying morphological operation

Fig. 3. The set of Haar-like features for AdaBoost

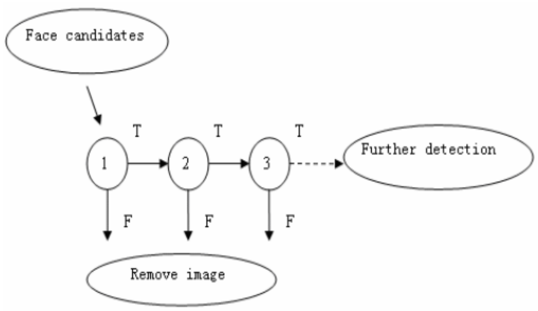


Fig. 4. Schematic depiction of a cascaded detector

After calculation of a huge number of features for each analysis window, the AdaBoost algorithm is used for combining a small number of these features to form an effective classifier.

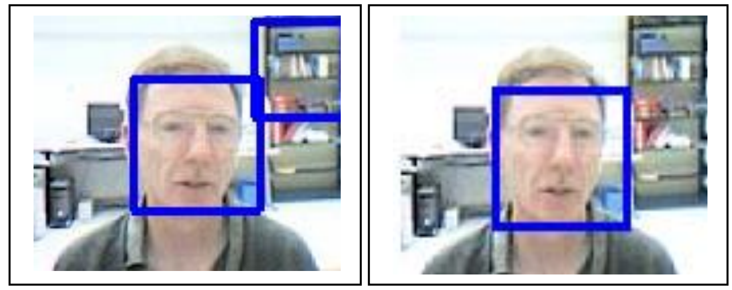
C. Skin color verification

As mentioned, Viola and Jones face detector has some substantial limitations, as a high false positives rate. In order to solve problem, Skin color post- filtering is used to verify if there are sufficient number of skin pixels in a face candidates sub-windows detected by VJFD; all sub-windows with a size smaller than Ω = 300 pixels are discarding. Verification with skin color filtering is illustrated in Figure.5.

IV. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed method, many experiments have been carried out using a variety of face datasets, which include samples video taken from TV, video-conference cameras, and standard video database NRC-IIT [1]. The NRC-IIT database, publicly available, contains 11 subjects, and for each subject there is a pair of low-resolution video clips. Each clip shows a face of a computer, user sitting in front of the monitor, who undergoes a variety set of movements, which cause changes in face illumination.

We also compare our proposed method with three methods. The first one [10] applies skin color based pre-filtering prior to the application of the VJFD. The entire image is scanned to retrieve face candidates by skin color segmentation and morphological operations, then the final face regions are obtained by applying the VJFD to these face candidates. The second method [13] uses skin color as a post-processing stage to the VJFD. The third method is the later detector reported in [07]. The comparison is based on the accuracy, the false positives rate, and the false negatives rate, which are defined below table1 [30].



(a) **(b)**

Fig. 5. Skin color verification (a) the face detection results of the VJFD [7] algorithm for image 29 of the video 00-1.avi are shown. (b) Face detection verified by skin color filtering. The false positive has been eliminated.

The results presented in table 1 were obtained using the video clips of subjects from the NRC-IIT database. The analysis of these results show clearly, that the accuracy of the VJFD increases when a skin color based pre-filtering is used, which means that this face detector gives *low false negatives rate* when it is not applied directly to the entire input image. This motivates the use of a skin color detector as a prior stage in our method. However, the skin color detector may miss some skin pixels, which results in false negatives.

Overall, it can be stated that **i**ntroducing saliency based visual attention in a face detection system improves significantly the accuracy by overcoming the skin color failure to detect skin regions, especially under illumination variations, caused by movement, and in the case of distant face detection in complex backgrounds.

It should be noticed that integrating visual attention based face detection to skin color filtering into the VJFD pre-processing does not increase the processing time on a parallel architecture computer, since the two processing may be run simultaneously. Furthermore, incorporating a prior pre-filtering stage to VJFD may reduce the processing time of this detector since it is then applied to only a portion of the entire image, the portion that

Table I
Comparisons of our results with other face detectors methods

	Accuracy	FNR	FPR
Viola & Jones [07]	79.46%	20.54%	34%
VJFD and Skin-post-filtering [13]	80.28%	19.72%	09.12%
Skin pre-filtering and VJFD [10]	82.10%	17.92%	25.87%
Proposed method	91.54%	07.83%	09.10%

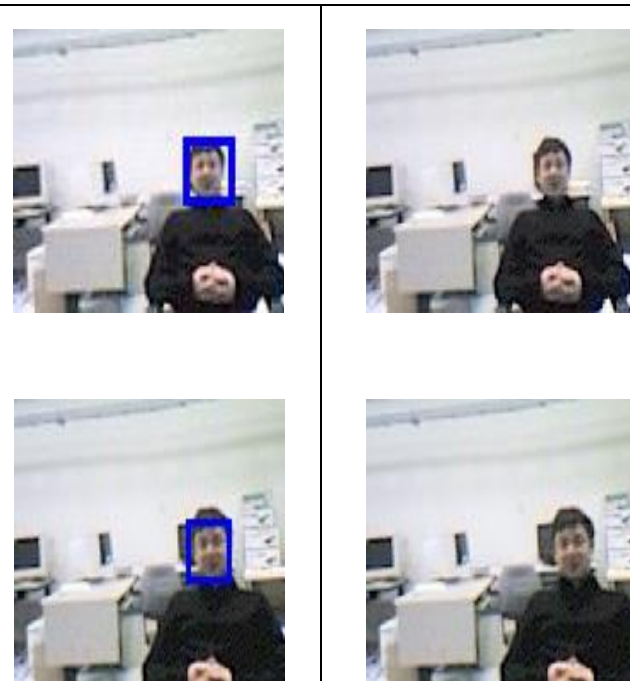
- False Positives Rate (**FPR**) = the ratio of the number of detected false positives to the total number of faces.
- False Negatives Rate (**FNR**) = the ratio of the number of false negatives to the total number of faces.
- Accuracy = 1-(FPR+FNR)/2

may reduce the processing time of this detector since it is then applied to only a portion of the entire image, the portion that contains face candidates provided by the pre-filtering stage.

V. Conclusion

This paper presents a novel face detection method which reduces the false negatives, and the false positives rates of the well known method Viola and Jones face detector. The proposed method applies visual attention based face segmentation and skin color based filtering, prior to the application of the VJFD, to reduce the false negatives rate. It then applies skin color post-filtering to eliminate false positives. The experimental results presented illustrate the effectiveness of this method, compared to some other methods proposed in the literature, especially the well known VJFD.

In a future work we will try to evaluate the complexity of the proposed method, in terms of computational load, and compare it to that of other methods.



(a) (b)

Fig. 6. test methods at different distance (a) results of our proposed method (b) skin color method and VJFD failed to detect face .

REFERENCES

[1] Yang M.H., Kriegman D., and Ahuja N., "**Detecting Faces in Images**: A Survey", **IEEE** Trans. Pattern Analysis Machine Intelligence, 24(1), 2002.

[2] Yunawen Wu, and Xueyi Ai, "Face Detection in Color Images Using AdaBoost Algorithm Based on Skin Color Information", First International Workshop on Knowledge Discovery and Data Mining, 2008, pp.339-342

[3] Kwok-Wai Wong, Kin-Man Lam, and Wan-Chi Siu, "An Efficient Algorithm for Human Face Detection and Facial Feature Extraction under Different Conditions", Pattern Recognition, Vol. 34,2001, pp.1993-2005.

[4] Ing-Sheen Hsieh, Kuo-Chin Fan, and Chiunhsiun Lin, "A Statistic Approach to the Detection of Human Faces in Color Nature Scene", Pattern Recognition, 35, 2002, pp.1583-1596.

[5] J. G. Wang and T. N. Tan, "A New Face Detection Method Based on Shape Information", IEEE Transactions on Pattern Recognition Letters, Vol. 21, 2000, pp. 463-471.

[6] Chiunhsiun Lin and Kuo-Chin Fan, "Triangle-based Approach to the Detection of Human face", Pattern Recognition, Vol. 34, No. 6, 2001, pp.1271-1283.

[7] Paul Viola and Michel J. Jones: "Rapid Object Detection Using a Boosted Cascade of Simple Features" , *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.511–518, 2001.

[8] Cha Zhang, and Zhengyou Zhang, "boosting-based face detection and adaptation, Synthesis", Lectures on Computer Vision (Book) by Cha Zhang, et al. 2010.

[9] Yan-Wen Wu, "Face Detection in Color Images Using AdaBoost Algorithm Based on Skin Color Information", Huazhong Normal Univ., Wuhan, Xue-Yi Ai.

[10] Z.S. Tabatabaie, R.W. Rahmat, N.I. Udzir and E.Kheirkhah, "A hybrid face detection system using combination of appearance-based and feature-based methods," *IJCSNS International Journal of Computer Science and Network Security*, vol.9, no.5, May 2009.

[11] M. Niazi, and S. Jafari, "Hybrid face detection in color images," *IJCSI International Journal of Computer Sciences Issues*, vol. 7, pp. 367–373, 2010.

[12] C.E. Erdem, S. Ulukaya, A. Karaali, A.T. Erdem, "Combining haar feature and skin color based classifiers for face detection," *IEEE 36th International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*, 2011.

[13] Shanshan Wang and Amr Abdel-Dayem, "Improved Viola-Jones Face Detector", Department of Mathematics and Computer Science, Laurentian University Sudbury, Canada.

[14] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods, pattern recognition", vol.40, no3, pp.1106-1122, 2007.

[15] V. Vezhnevets, V. Sazonov, and Andreeva A., "A survey on pixel-based skin color detection techniques," *Proc. Graphicon*, 2003.

[16] Deepak Ghimire and Joonwhoan Lee, "A Robust Face Detection Method Based on Skin Color and Edges," J. Inf. Process Syst., Vol.9, No.1, March 2013.

[17] E.Mostafa, A.El-Barkouky,H.Rara, A. Farag, "Rejecting Pseudo-Faces using the Likelihood of Facial Features and Skin",CVIP Lab, University of Louisville, KY, USA, International Conference on Biometrics: Theory, Applications and Systems (BTAS), IEEE, 2012.

[18] Zhang, L., Tong, M. H., Marks, T. K., Shan, H., and Cottrell, "A Bayesian framework for saliency using natural statistics", Journal of Vision, 8(7), 1-20.

[19] V. Navalpakkam and L. Itti, "An integrated model of top-down and bottom-up attention for optimal object detection", in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, Jun 2006, pp. 2049–2056.

[20] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in Advances in Neural Information Processing Systems 19, B. Scholkopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press, 2007.

[21] L. Zhang, M. H. Tong, and G. W. Cottrell, "Information attracts attention: A probabilistic account of the cross-race advantage in visual search", in Proceedings of the 29th Annual Cognitive Science Conference, 2007.

[22] Nicholas J. Butko, Lingyun Zhang, Garrison W. Cottrell, and Javier R. Movellan , "Visual Saliency Model for Robot Cameras", Proceedings of the IEEE International Conference on Robotics and Automation, May, p.2398 -2403, (2008). ICRA '08..

[23] W. Kienzle, F. A. Wichmann, B. Sch¨olkopf, and M. Franz, "A nonparametric approach to bottom-up visual saliency," in Advances in Neural Information Processing Systems 19, B. Scholkopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press, 2007.

[24] Moran Cerf , E. Paxon Frady , Christof Koch, "Face and text attract gaze independent of the task Experimental data and computer model", Computation and Neural Systems 2010.

[25] A.Shahab, F.Shafait, A.Dengel " How Salient is Scene Text? ", IEEE Conference on Document Analysis Systems (DAS), 2012 10th IAPR .

[26] L. Elazary and L. Itti, "A Bayesian model for efficient visual search and recognition,",Vision Research, vol. 50, no. 14, pp. 1338–1352, 2010.

[27] N. D. B. Bruce and J. K. Tsotsos, "Saliency based on information maximization," in Advances in Neural Information Processing Systems 18, Y. Weiss, B. Sch¨olkopf, and J. Platt, Eds. Cambridge, MA: MIT Press, 2006, pp. 155–162.

[28] Xiaodi Hou, Liqing Zhang, "Saliency Detection: A Spectral Residual Approach," *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp.1-8.

[29] B. Funt, K. Barnard, and L. Martin, "Is machine color constancy good enough?," in European Conference on Computer Vision, 1998, pp. 445–459.

[30] Rank Y. Shih , "Image Processing and Pattern Recognition: Fundamentals and Techniques", chapter 10", page 367.

[31] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254–1259, Nov 1998.