Hybrid Face Detection System using Combination of Viola - Jones Method and Skin Detection

Amr El Maghraby
Ph.D Student at Zagazig Univ.

Mahmoud Abdalla
Prof. at Zagazig Univ.

Othman Enany
Ph.D at Zagazig Univ.

Mohamed Y. El Nahas
Prof. El Azhar Univ.

Faculty of Engineering, Computers and Systems Engineering Dept., Zagazig University, Egypt

ABSTRACT

In this paper, a fast, reliable automatic human face and facial feature detection is one of the initial and most important steps of face analysis and face recognition systems for the purpose of localizing and extracting the face region from the background. This paper presents a Crossed Face Detection Method that instantly detects low resolution faces in still images or video frames. Experimental results evaluated various face detection methods, providing complete solution for image based face detection with higher accuracy, showing that the present method efficiently decreased false positive rate and subsequently increased accuracy of face detection system in still images or video frames especially in complex backgrounds.

General Terms

Image processing, Face detection, Algorithms

Keywords

Face detection, Videos frames, Viola- Jones, Skin detection, Skin color classification.

1. INTRODUCTION

Face detection is an easy visual task for human vision, however; this task is not easy and is considered to be a challenge for any human computer interaction approach based on computer vision because it has a high degrees of variability in its appearance. How can computers detect multiple human faces present in an image or a video with complex background? That is the problem. The solution to this problem involves segmentation, extraction, and verification of faces and possibly facial features from complex background.

Computer vision domain has various applications [1] such as Face Recognition, Face localization, Face Tracking, Facial Expression Recognition, Passport Control, Visa Control, Personal Identification Control, Video Surveillance, Content-Based Image and Video Retrieval, Video Conferencing, Intelligent Human Computer Interfaces and Smart Home Applications. Challenges faced by face detection algorithms often involve the following: 1- Presence of facial features such as beards, moustaches and glasses. 2- Facial expressions and occlusion of faces like surprised or crying. 3-Illumination and poor lighting conditions such as in video surveillance cameras image quality and size of image as in passport control or visa control. 4-Complex backgrounds also makes it extremely hard to detect faces [14]. Face detection techniques have been researched for years and much progress has been proposed in literature. The most five known algorithms [2] for face detection are: Principle Component Analysis (PCA),

Linear Discriminator Analysis (LDA), Skin Color, Wavelet and Artificial Neural Networks . Most of the face detection techniques focus on detecting frontal faces with good lighting conditions in images or videos. Various face detection algorithms have been proposed. These numerous methods could be grouped into two main approaches:

a) Feature-based techniques: The feature based techniques extract local regions of interest features (eyes, nose , etc...) from the images and identify the corresponding features in each image of the sequence [14].

b) Image based techniques: Which uses classifiers trained statically with a given set of samples to determine the similarity between an image and at least one training sample. The classifier is then scanned through the whole image to detect faces.

2. Viola-Jones object detection framework

The Viola-Jones [3][4] object detection framework proposed in 2001 was the first object detection framework to provide competitive object detection rates in real-time. It could detect faces in an instant and robust manner with high detection rates. The Viola-Jones face detector analyzes a given sub-window using features consisting of two or more rectangles presenting: The different types of features[3] (see Figure 1).

Figure 1: The different types of features

Although it can be trained to detect a variety of object classes, it was motivated primarily by the problem of face detection. This algorithm is implemented in Open CV as CV Hear Detect Objects [5]. First, a classifier (namely a cascade of boosted classifiers working with Hear-like features) is trained with a few hundred sample views of a particular object (i.e., a face or a car), called positive examples, that are scaled to the same size (say, 20x20), and negative examples - arbitrary images of the same size.

After a classifier is trained, it can be applied to a region of interests (of the same size as used during the training) in an input image. The classifier outputs "1" if the region is likely to show the object (i.e., face or upper body); otherwise it gives "0". To search for the object in the whole image one can move the search window across the image and check every location

using the classifier. The classifier is designed so that it can be easily "resized" in order to be able to find the objects of interest at different sizes, which is more efficient than resizing the image itself. In order to find an object of an unknown size in the image scan procedure should be performed several times on different scales basis.

The Viola - Jones method contains three techniques:

1. Integral image for feature extraction
2. AdaBoost [6][7]for face detection
3. Cascade classifiers [9]

2.1 Integral Image for Feature Extraction Techniques

The first step of the Viola-Jones object detection framework is to turn the input image into an integral image defined as two-dimensional lookup tables (see Figure 2) to a matrix with same size as the original image. The integral image at location of x,y = sum of all pixel values above and to the left of (x,y). Each element of the integral image contains the sum of all pixels located on the upper-left region of the original image (in relation to the element's position). This allows computing sum of rectangular areas rapidly at any position or scale by using only four values. These values are the pixels in the integral image that co-exist with the corners of the rectangle within the input image (see Figure 2).

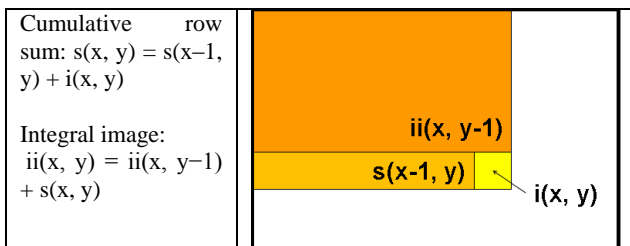| Cumulative row sum: s(x, y) = s(x−1, y) + i(x, y) Integral image: ii(x, y) = ii(x, y−1) + s(x, y)	

Figure 2: Computing the integral image [16]

A window of the target size is moved over the integral images, and for each subsection of the image the Haar-like feature [9] is calculated. This difference is then compared to a learned threshold that separates non-objects from objects. To calculate the Rectangle Feature value (f) of the box enclosed by the dotted line (see Figure 3).
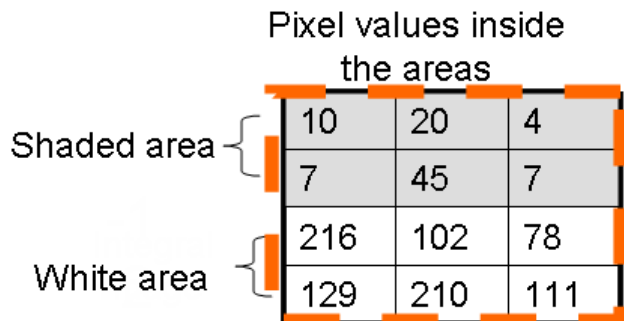
Pixel values inside the areas

10	20	4
7	45	7
216	102	78
129	210	111

Shaded area
White area

Figure 3- Calculation of Rectangle Feature Value [16]

f= $\sum$ (pixels in white area) – $\sum$ (pixels in shaded area)
f= (216+102+78+129+210+110) -
    (10+20+4+7+45+9) = 720
If f >threshold,
Feature=+1 ➔(object)
Else

Feature=-1 ➔(non object)

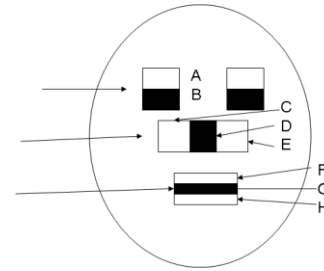Face consists of many features , different sizes, polarity and aspect ratios (see Figure 3).

Figure 3: Example of Face Features

These features could be considered as rectangular face features.
Two eyes= (Area_A - Area_B)
Nose =(Area_C+ Area_E- Area_D)
Mouth =(Area_F+ Area_H -Area_G)
The eye-area (shaded area) is dark; the nose-area (white area) is bright. So f is large, hence it is face

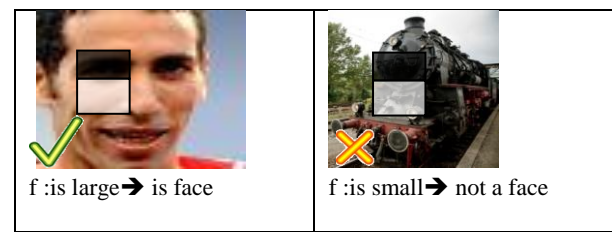| f :is large ➔ is face	f :is small➔ not a face

Figure 4: Detect Face and Non Face by Rectangle Feature Value

The value of any given feature is as follows:
The sum of pixels within clear rectangles subtracted (-) from the sum of pixels within shaded rectangles.

2.2 AdaBoost for Face Detection

AdaBoost (adaptive boosting) is a machine learning algorithm [6] which can be used for classification or regression. It combines many small weak classifiers to become a strong classifier, using only a training set and a weak learning algorithm, AdaBoost is called adaptive because it uses multiple iterations to generate a single strong learner. AdaBoost creates the strong learner by repeatedly adding weak learners. During each round of training, a new weak learner is added to the ensemble and a weighting vector is adjusted to focus on examples that were misclassified in previous rounds. In Viola-Jones frame work, Haar-like features [9] are used for rapid objects detection and supports the trained classifiers, Haar-like features are the input to the basic classifiers, and are calculated as described below. The basic classifiers are decision-tree classifiers with at least 2 leaves. Where x is a 24*24 pixels sub-window, f is the applied feature, p the polarity and θ the threshold that decides whether x should be classified as a positive (a face) or a negative (a non-face)

$$h_t(x) = \begin{cases} 1 & \text{if } p_t f_t(x) < p_t \theta_t \\ 0 & \text{otherwise} \end{cases}$$

value of rectangle feature
window
threshold
$p_t=$ polarity{+1,-1}

2.3 Cascade Classifier

The word "cascade" in the classifier name means that the resultant classifier consists of several simpler classifiers (stages) of multiple filters to detect the Haar-like features [9] that are applied subsequently to a region of interest until at some stage the candidate is rejected or all the stages are passed. In order to greatly improve computational efficiency (high speed and high accuracy) and reduces the false positive rates. Viola-Jones uses cascaded classifiers composed of stages each containing a strong classifier. Each time the sliding window shifts, the new region within the sliding window will go through the cascade classifier stage-by-stage. If the input region fails to pass the threshold of a stage, the cascade classifier will immediately reject the region as a face. If a region passes all stages successfully, it will be classified as a candidate of face, which may be refined by further processing. The job of each stage is to determine whether a given sub-window is definitely not a face or maybe a face. When a sub-window is classified to be a non-face by a given stage it becomes immediately a discarded figure.(see fig.5). Conversely a sub-window classified as a maybe-face is passed on to the next stage in the cascade. The concept is illustrated with multi stages in 6.

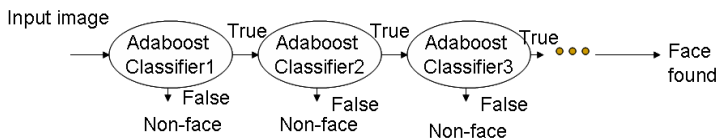Figure 5: The sliding window shifts.

Figure 6: The cascade classifier.

3. Skin Color Detection

Skin detection in color images and videos is a very efficient way to locate skin-colored pixels. Skin color is a distinguishing feature of human faces. In a controlled background environment, skin detection can be sufficient to locate faces in images. As color processing is much faster than processing other facial features, it can be used as a preliminary process for other face detection techniques [11]. Skin detection has also been used to locate body limbs, such as hands, as a part of hand segmentation and tracking systems, e.g., [12]. However, many objects in the real world have skin-tone colors, such as some kinds of leather, sand, wood, fur, etc., which can be mistakenly detected by a skin detector. Therefore, skin detection can be very useful in finding human faces and hands in controlled environments where the background is guaranteed not to contain skin-tone colors. Since skin detection depends on locating skin-colored pixels, its use is limited to color images, i.e., it is not useful with gray-scale, infrared, or other types of image modalities that do not contain color information.

Several computer vision approaches have been developed for skin detection. A skin detector typically transforms a given pixel into an appropriate color space and then uses a skin classifier to label the pixel whether it is a skin or a non-skin pixel. A skin classifier defines a decision boundary of the skin color class in the color space based on a training database of skin-colored pixels. Different classes of color spaces are the orthogonal color spaces used in TV transmission. This includes YUV, YIQ, and YCbCr. YIQ which is used in NTSC TV broadcasting while YCbCr is used in JPEG image compression and MPEG video compression. One advantage of using these color spaces is that most video media are already encoded using these color spaces. Transforming from RGB into any of these spaces is a straight forward linear transformation [13].

The proposed framework is based on Transforming from RGB to HSV color space and YCbCr chrominance space.

In this section image processing techniques and different operations for different regions on the same image are applied as detail experimental skin detection
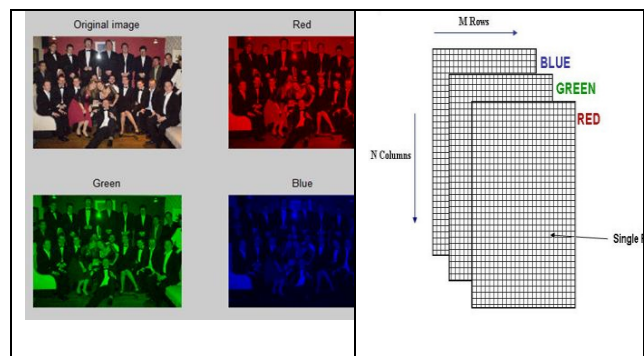

Figure 7: Converting color image into RGB color space.

3.1 Building Skin Model

First an image from database which having default RGB color space is taken. Then RGB to HSV color space conversion is performed so that threshold for skin color region using HSV could be found. HSV-type color spaces are deformations of the RGB color cube and they can be mapped from the RGB space via a nonlinear transformation. One of the advantages of these color spaces in skin detection is that they allow users to intuitively specify the boundary of the skin color class in terms of the hue and saturation.
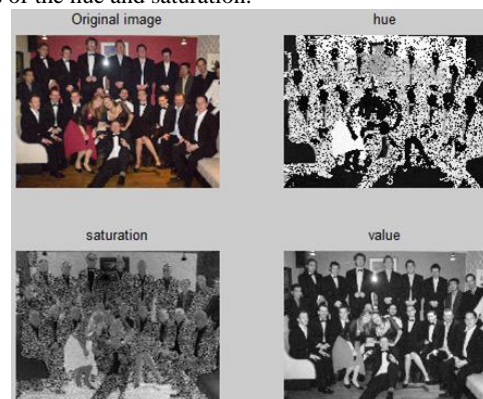

Figure 8: Converting RGB space into HSV space

Similarly, RGB to YCbCr color space conversion is performed to find out threshold for skin region using the following equation.

$$Y = 0.257R + 0.504G + 0.098B + 16$$
$$Cb = -0.148R - 0.291G + 0.439B + 128$$

$$Cr = 0.439R - 0.368G - 0.071B + 128$$

In next step, combined all above color spaces based on the basic idea of Venn diagram and finally mask the skin color region.
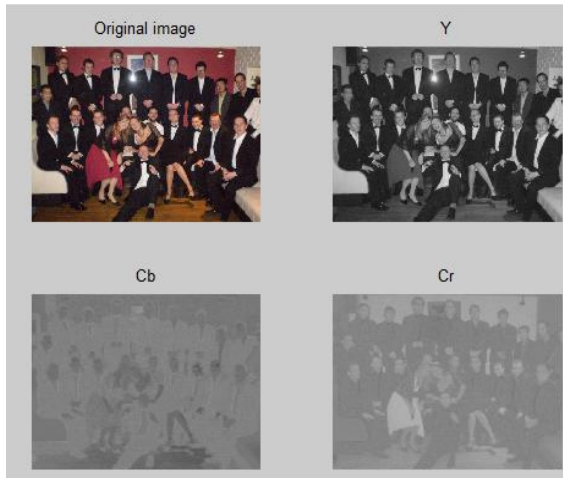

Figure 9: Converting RGB space into YCrCb space

3.2 Skin Segmentation

The first stage is to transform the image to a skin-likelihood image. This involves transforming every pixel from RGB representation to chroma representation and determining the likelihood value based on the equation $140< Cr < 165$ &
$140 <Cb< 195$ a region of orange to red to pink in red-difference and blue-difference channels $0.01 < Hue < 0.1$ this means hue is basically reddish [14]. Since the skin regions are brighter than the other parts of the images, the skin regions can be segmented from the rest of the image through an optimal threshold, all pixel values which have likelihood values higher than threshold are set to 1 and the rest of pixels are set to 0.

3.3 Area opening operation

The objective of this step is to remove, from a resulting binary image, all connected components objects that have fewer than threshold pixels, producing another binary image.

Figure 10: Before removing small connected pixels

Figure11: After removing small connected pixels

3.4 Morphological Operations

Now binary image with 1's representing skin pixels and 0's representing non-skin pixels is obtained. Then morphological operations such as filling, erosion and dilation are applied in order to separate the skin areas which are loosely connected. Morphological closing is applied primarily to the binary image then; aggressive morphological erosion is applied by using structuring element of disk size 10. Erosion operation examines the value of a pixel and its neighbors and sets the output value equal to the minimum of the input pixel values. Morphological dilation is applied to grow the binary skin areas which are lost due to aggressive erosion in previous steps by examining the same pixels and outputs the maximum of these pixels.
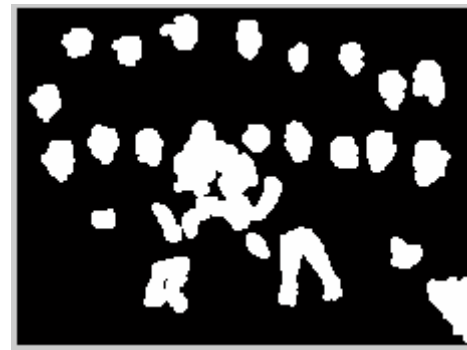

Figure 12: Removal of small connected pixels.

3.5 Region Labeling

The process of determining how many regions are found in a binary image and are done by labeling such regions. A label is an integer value. In order to determine the labeling of a pixel, an 8-connected neighborhood (i.e., all the neighbors of a pixel) are used. If any of the neighbors had a label, the current pixel is labeled. If not, a new label is used. At the end, the number of labels are counted and this will be the number of regions in the segmented image.

Figure 13: Region labeling.

3.6 Template Matching

This is the final stage of face detection where cross correlation between template face and grayscale region. The gray scale regions containing faces are extracted by multiplying the binary region with grayscale original image (see Figure 14). The final result (see Figure 15).

Figure 14: Template matching result.

Figure 15: Final result.

4. Proposed Face Detection Framework

The proposed method illustrated in this paper on the face detection task, extends the cascade object detector framework proposed by Viola-Jones. [3] It's main objective is to detect rotated and profiled faces in still images or videos, especially in complex background images, by using Viola Jones upper body model to detect near-frontal upper-bodies as a region of interest. This is the primary detector where the high probability of finding the face instead of searching the entire image. In order to find an accurate face in that region of interest, Viola-Jones [4] face detector is used as a secondary detector to increase accuracy and reduces false negatives. Third detector pixel-based skin detection methods are applied on the region of interest which is not detecting a face using the secondary detector. The third detector classifies each pixel as skin or non-skin individually and independently from its neighbors and combines it with Viola - Jones upper body detection. This improves the performance of face detection systems in terms of increasing the face detection speed and decreasing false positive rate. The primary, secondary and third detectors are combined by this release and a single homogeneous set of face bounding-boxes are returned. The algorithm can detect a face in video as usual in each image. Once a face is detected, a mixture of face detection and tracking from the point detected is used.
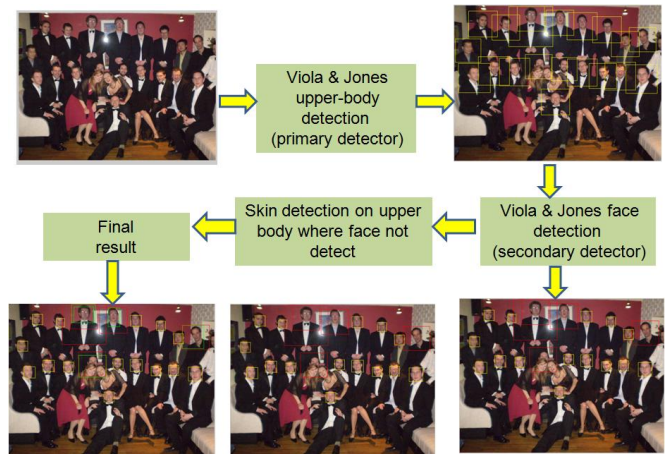
Figure 16: Demonstrates overall schema of the proposed face detection system.

5.Experimental Result

In this section a detailed experimental comparison of the above stated approaches are being presented. Fifty test images were obtained in different lighting conditions and complex backgrounds with comparison of three algorithms on same images. The first algorithm used was Viola - Jones face detection[4], the second algorithmic approach was described by Zahra Sadri[15] who used skin color detection in the input image first and then applied Viola - Jones algorithm for facial detection (see Figure 17), The previously mentioned approach are completely different from the proposed algorithm.

Figure 17: Face Detection System using combination of Appearance-based and Feature-based methods [15]

The below figures are screen shots for the experimental comparison of all three methods presented. We used 50 color images containing many faces with various complex backgrounds. The resulting detector automatically returned bounding-boxes fitting detecting faces and automatically counting number of faces appearing in images.

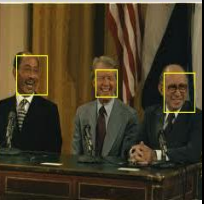| Viola& Jones face detection	Apply Viola& Jones face detection on skin region	Proposed algorithm

16 face detected	13 face detected	20face detected
2 face detected	0 face detected	3 face detected
0 face detected	0face detected	1 face detected
0 face detected	0 face detected	1 face detected

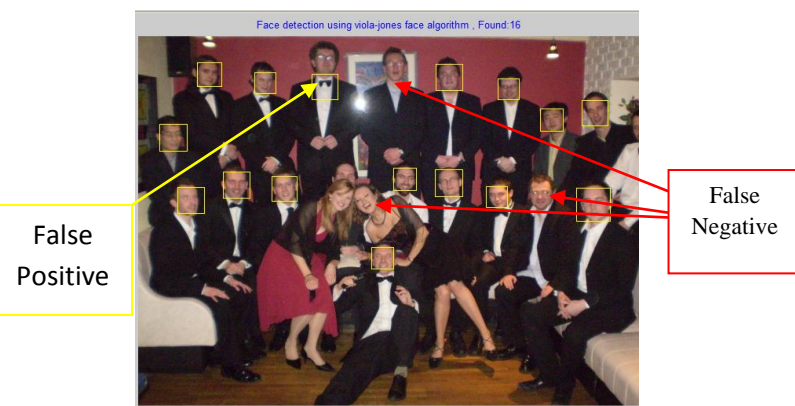Figure 18: comparison of face detection between three methods

Figure 19: Faces detected by Viola - Jones face detector on an experimental sample image

Figure 20: Proposed algorithm results Draw automatically the returned bounding box around the face detected

The hereunder figures demonstrated the statistical results of the experimental results on 50 test images representing faces at different imaging conditions. The detector applied the 3 different methods.

1- Viola - Jones face detection system
2- Apply Viola - Jones face detection on skin region
3- Proposed algorithm

Experimental results showed that the proposed method improved face detection especially for people in different poses and difficult backgrounds. The proposed algorithm clearly reduced false negative rates and detected face poses from the images at different imaging resolution and different lighting conditions.
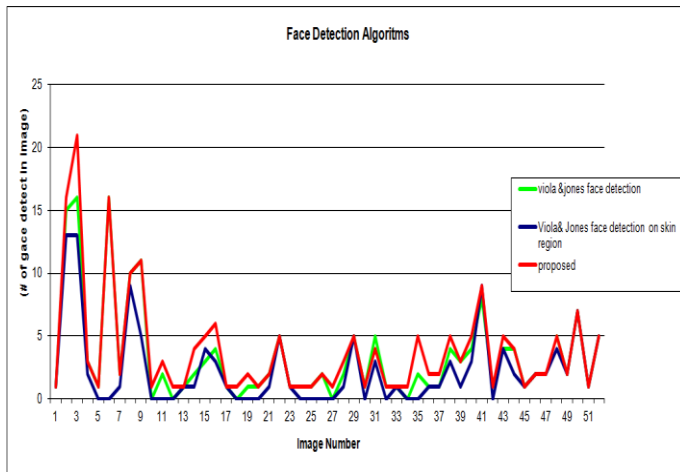

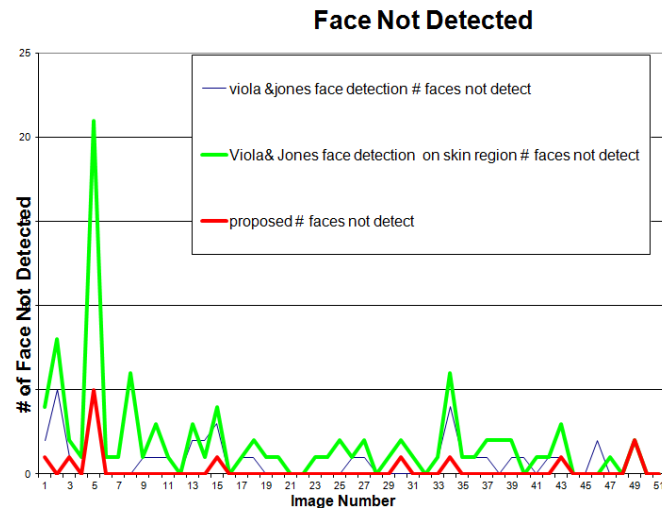Figure 23: Comparison of Face Detection rate for three methods.

Figure 21: comparison for number of face detected in 50 different images for three methods

Table 1: Comparison of Face Detection Accuracy for three methods.

Criteria	Viola& Jones face detection	Viola& Jones face detection on skin region	Proposed algorithm
False Positive Rate	3.738318	20.09345794	7.943925
False Negative Rate	20.09346	55.14018692	6.074766
Accuracy	76.16822	24.76635514	85.98131

Figure 22: Face Not Detected

The accuracy is obtained by using the following equation:

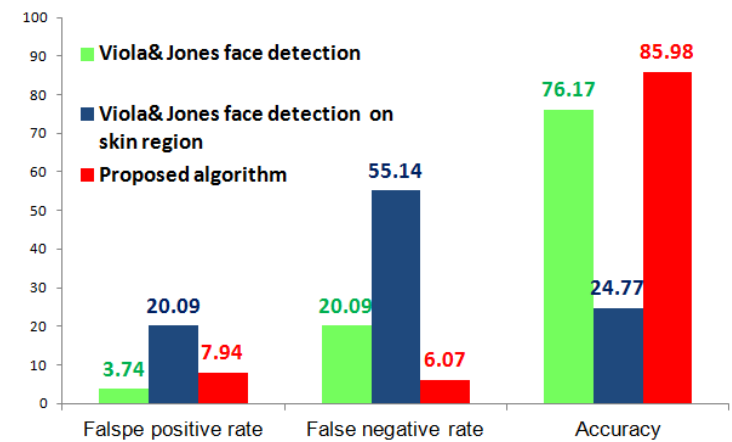$$\% \text{ Accuracy} = 100 - (\text{False positive Rate} + \text{False negative Rate})$$

Figure 24: Comparison of Face Detection Accuracy for the three methods.

6. CONCLUSION AND FUTURE SCOPE

In this paper, a novel face detection algorithm using combination of face detector methods is presented to improve face detection that automatically detects face in complex background. The proposed algorithm start by finding region of interest using Viola - Jones upper body as first detector then apply Viola - Jones face on region of interest as second detector and finally apply skin detector as third Feature-based. The system was implemented using Mat-lab Environment. The proposed method can process different kinds of images, and under different lighting conditions. The experimental results showed that our new approach was able to achieve a higher detection rate than any of the 2 methods mentioned prior, and clearly improved Viola - Jones face detection accuracy and decreasing false negative rates. This research work was initiated as a part of research project for Human Actions Detection In Content-based Video Retrieval System. In the future this algorithm will be a part of a system which will identify human presence in video.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] http://en.wikipedia.org/wiki/Face_detection

[2] N.Ismail," Review of Existing Algorithms for Face Detection and Recognition," 8th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics,December 14-16,2009,pp.30-39.

[3] Viola, P. and M. J. Jones (2004). "Robust Real-Time Face Detection." International Journal of Computer Vision 57(2): 137-154.

[4] Viola, P. and M. Jones (2001). Rapid Object Detection Using a Boosted Cascade of Simple Features, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition.

[5] http://opencv.org/

[6] K. T. Talele, S. Kadam, A. Tikare, Efficient Face Detection using Adaboost, "IJCA Proc on International Conference in Computational Intelligence", 2012

[7] Cristinacce, D. and Cootes, T. Facial feature detection using AdaBoost with shape constraints. British Machine Vision Conference, 2003.Singh, S. K., Chauhan, D. S., Vatsa, M. and Singh, R. (2003). "A Robust Skin Color Based Face Detection Algorithm." TAMKANG.

[8] Papageorgiou, Oren and Poggio, "A general framework for object detection", International Conference on Computer Vision, 1998.

[9] Phillip I.W.,Dr. John F. FACIAL FEATURE DETECTION USING HAAR CLASSIFIERS,JCSC 21, 4 (April 2006)

[10] Bernhard Fink, K.G., Matts, P.J.: Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. Evolution and Human Behavior 27(6) (2006) 433–442

[11] Senior, A., Hsu, R.L., Mottaleb, M.A., Jain, A.K.: Face detection in color images. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI) 24(5) (2002) 696–706

[12] Imagawa, K., Lu, S., Igi, S.: Color-based hands tracking system for sign language recognition in: FG '98: Proceedings of the 3rd. International Conference on Face & Gesture Recognition, Washington, DC, USA, IEEE Computer Society (1998) 462

[13] Burger, W., Burge, M.: Digital Image Processing, an Algorithmic Introduction Using Java. Springer (2008)

[14] Yang, M., Kriegman, D., Ahuja, N.: Detecting faces in images: A survey. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI) 24(1) (2002) 34–58

[15] Zahra S. T., Rahmita W. R., Nur I. B.\ and Esmaeil K." A Hybrid Face Detection System using combination of Appearance-based and Feature-based methods." International Journal of Computer Science and Network Security, VOL.9 No.5, May 2009

[16] [Lazebnik09]www.cs.unc.edu/~lazebnik/spring09/lec23_face_detection.ppt