A Novel Approch For Real Time Face Detection Over A Full HD Video

Anokhi Gandhi
121005
Kashyap Patel
121021
Shashwat Sanghavi
121049
Sohum Shah
121056
Team: DOBA
Information and Communication Technology
Institute of Engineering & Technology, Ahmedabad University (IET-AU)
Ahmedabad, India

***Abstract*—Human faces are one of the most important resource of knowledge for an intelligent vision based HCI system. An intelligent machine must learn how to analyze human behavior with full automation. Faces primarily define human behavior. Face recognition has received much attention and is widely used in numerous applications like computer vision communication and automatic access control system. Face detection is essentially the first part of automatic face recognition, however not so straightforward. It faces challenges imposed by image appearance, pose variation, occlusion, image/face orientation, illuminating condition, background and facial expression. Several techniques have been developed and implemented to detect faces from an image. This paper throws light on a technique to detect and track faces in any given high definition video.**

***Index Terms*—Real time face-detection, Face tracking.**

I. Introduction

Face detection is a method that detects faces in a digital image. Many methods have been proposed to resolve the challenges imposed by an image containing a face. Certain examples of these methods include: template-matching methods: computing the correlation of an input image with a standard face pattern; feature-invariant method: detecting face features like eyes, nose, ears etc.; appearance-based methods: face detection with eigenface; neural network: identify the neural network of eyes-nose-mouth; nevertheless, implementing these methods independently is also a big challenge.

This paper proposes an algorithm for face detection and tracking in a given video. A video is a collection of images, wherein the images are displayed and put off at a high rate (frames per second). **The task is to first detect faces from a video with a frame rate of 30 fps, resolution of 1280x720 pixels. The algorithm should make sure to detect multiple faces in one frame, and also ignore faces of human-like cartoon characters and puppets.** Also, the processing of images should be done in **real time** so that the output video frame rate does not decrease than the original. An added advantage would be detecting of side-faces or unclear faces.

Face detection is done via the algorithm proposed by Viola and Jones[1],[4] and the face tracking is done via the CAMSHIFT Algorithm [2]. Viola and Jones algorithm is a rapid face detection algorithm, that is capable of processing images extremely rapidly while achieving high detection rates. Also, It detects face features more efficiently. There are three main parts in the Viola-Jones algorithm: firstly using Integral Images which allows a quick computation of features; second- A classifier using AdaBoost; third- method to combine classifiers in a cascade. Viola-Jones algorithm is explained in detail in later sections.

CAMSHIFT[2],[7] stands for Continuously Adaptive Mean Shift based on the mean shift algorithm. CAMshift uses the Hue channel to trace objects since by using the Hue channel based on HSV color model, objects with different colors can be recognized. Based on the color information, CAMshift tracks objects faster and consumes relatively little CPU resources. CAMshift is described in detail in following sections.

A. Literature review

Much work has been done to detect and track faces which have a wide area of applications. Below provided

is the brief explanation of 3 techniques studied for face detection and face tracking.

Viola and Jones face detection is a robust and fairly real time face detection technique [1],[4] which is based on training and learning principle. It uses **Haar features** to train itself and distinguish among faces and non faces. Based on the training set of faces it selects which features best detects a face and the features are now known as classifiers. This method is known as **Adaptive Boosting**. The image is then divided into windows and selected features are applied to individual windows. Since applying all features over the entire image involves complex and huge calculations the idea of **Cascading** is used. It divides the classifiers into sets and these sets are then applied to window. If these set of classifiers output a threshold more than decided, next set of classifiers are applied to the window; if it does not meet the threshold, the window is discarded and next window is analysed. These significantly reduces the number of calculations. Viola Jones face detection algorithm is said to be real time with the ability to process at a speed of 15 [1] fps on the machine with specific hardware.

Eigen Face detection [11] is face recognition method based on the Principle Component Analysis method.An important feature of PCA is that one can reconstruct any original image from the training set by combining the eigenfaces. This method can be applied only on gray images. The initial face images are normalized to line up eyes and mouth and resampled at same pixel resolution. The recognition process consists of initialization during which eigenface basis gets established through principle component analysis method; face classification is carried out where a new image is projected into face space. The resultant image according to the weight patterns is categorized as face or non face.

CAMSHIFT [7] is an object tracking method which tracks a particular given object. A search window is used which will try to find and follow the tracked object according to its position found in the previous frame. The histogram of the tracked object is calculated and now the search window moves towards the centre of mass accordingly. If target is met, Recalculate the centre of mass and set that window as new tracked object. This process gets repeated with all frames. Basically it shifts the tracked object according to the tracked object according to the mean.

II. Proposed algorithm

Viola Jones, Eigenface algorithm and CAMSHIFT have their own advantages as well as disadvantages. Viola & Jones can detect front face precisely but it takes a sufficiently long time to identify the face such that it cannot be used for real time. Eigenface algorithm, as well, is not feasible for real time face detection as one needs to take projection of the input image over principal components and this projection involves matrix multiplications which is computationally much more complex than Viola & Jones. On the other hand, CAMSHIFT is used to track objects. This algorithm can track object in a high definition video with more than 45 fps depending on the processor specifications. CAMSHIFT cannot detect faces.

A. Our approach

Considering the advantages and disadvantages of Viola & Jones and CAMSHIFT algorithm, this paper proposes an approach in which both face detection and CAMSHIFT algorithm are working in accordance with each other. The algorithm proposed in this paper takes a high definition video. First it finds for the face in a video. Once face(s) are found, they are stored in a list. Now each face stored in this list will be tracked using CAMSHIFT algorithm for 29 (can be varied depending on the video) subsequent frames. The reason for taking 29 frames is- given video is of 30 fps. Now in real world a human is not fast enough to move in/out of a given frame within a second. So detecting face for 1 frame and tracking face for rest 29 frames will not affect the accuracy of face detection. Here the selection of 29 fps depends on the frame rate of the video. Proposed algorithm is described in Algorithm 1.

Here face detection is done through Viola and Jones algorithm. For this one frame, we will compress the size of the frame to lower its dimensions and later the results(face detection box) will be expanded in order to match the original sized frame.The compression of the frame depends on the resolution of input video. It is approximately compressed by 5.In the case where no face is identified within a frame, this algorithm will be applied for identifying face(s) in subsequent frames until a face is found.

After detecting the face, instead of setting ROI as whole face, we will take only some part of the forehead. The reason of doing so is CAMSHIFT algorithm depends on the mean of the ROI. Viola jones returns the face with some background so mean of that ROI will be different than the actual mean of face. Now, actual mean of face will be same as mean of any skin image, so if we give only forehead as the ROI – for the first ROI – then, mean will be pure mean of skin which will make the algorithm

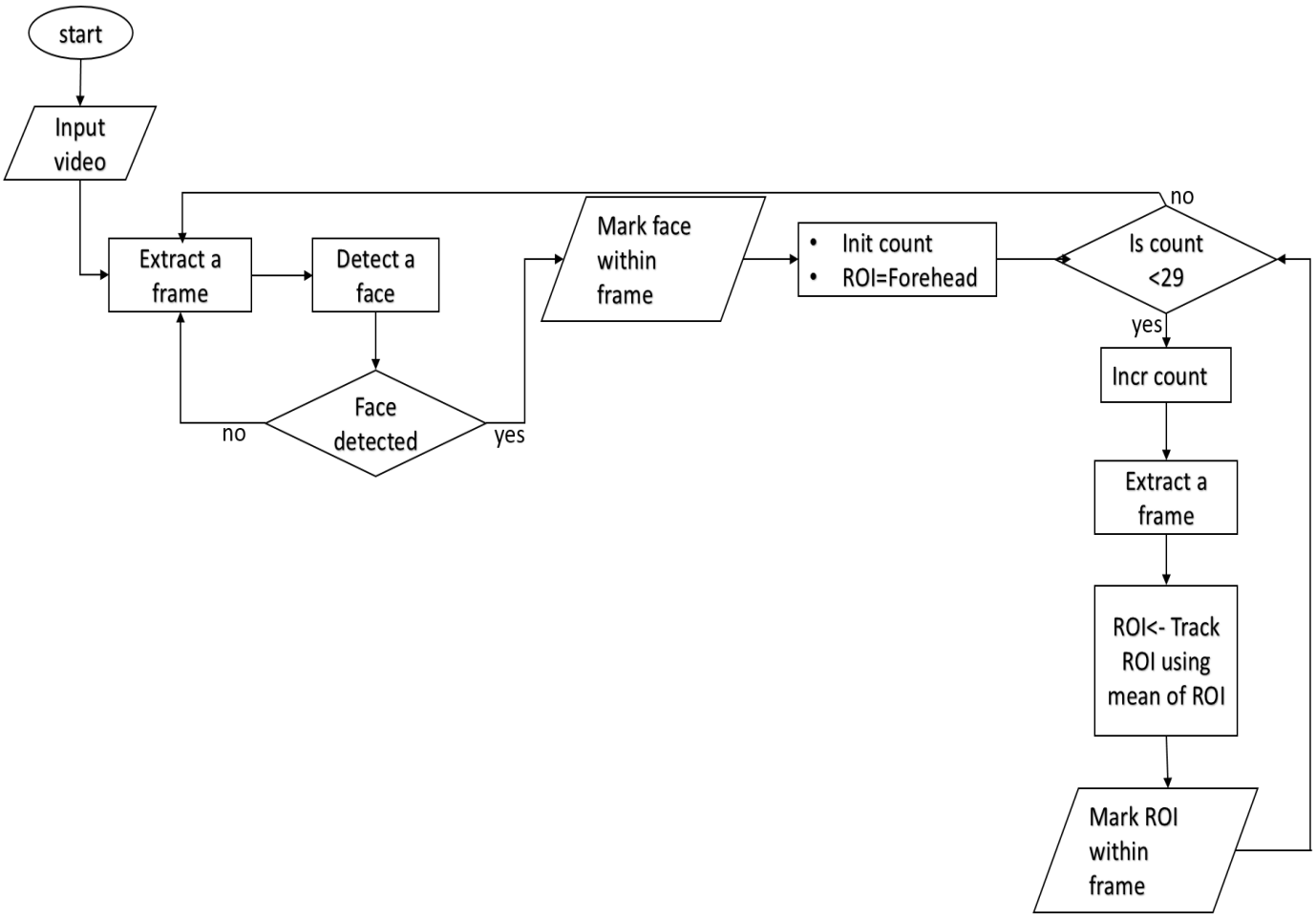

Fig. 1. Block diagram of proposed algorithm

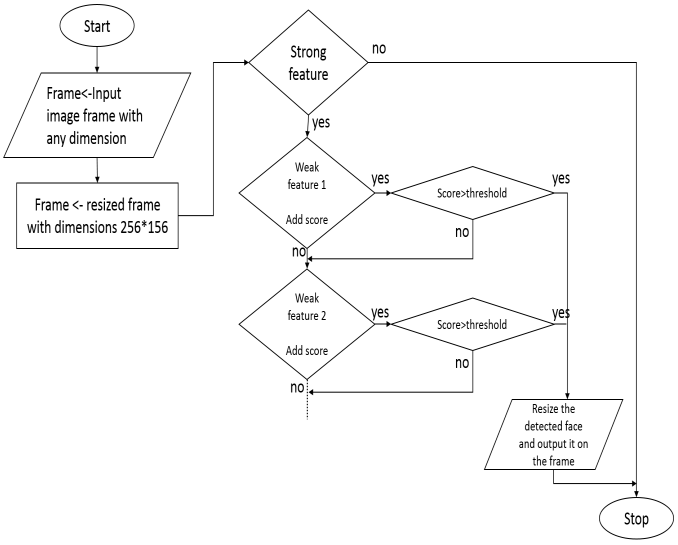

Fig. 2. Block diagram of Viola & Jones algorithm for face detection [10]

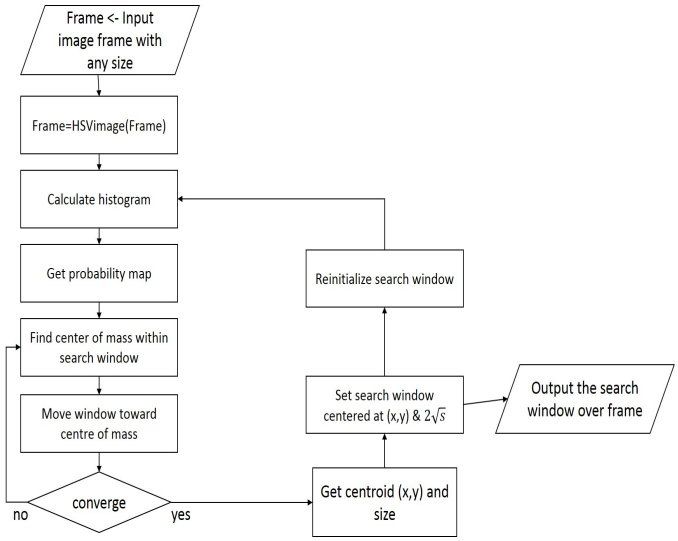

Fig. 3. Block diagram of CAMSHIFT algorithm for object tracking [9]

Data: video
Result: faces
1. Input a video.
2. Extract a frame.
3. **for** *frame is not NULL* **do**
   3.1 Resize the frame to 1/5th size.
   3.2 **for** *face is not identified* **do**
      3.2.1 Apply Viola Jones algorithm to this frame in order to identify the face.
   end
   3.3 Display the face over the frame.
   3.4 Set ROI(s) as forehead of the detected face(s).
   3.5 **for** *For subsequent 29 frames* **do**
      3.5.1 Extract a frame.
      3.5.2 Track ROI using CAMSHIFT.
      3.5.3 ROI¡- New ROI
      3.5.4 Display ROI over the frame.
   end
end

Algorithm 1: Real Time Face Detect and Track (RTFDT)

more precise in terms of locating the face in the frame.

Figure 1 shows the block diagram of Our algorithm. , Figure 2 and Figure 3 show the algorithm for detecting and tracking the face respectively. These are block diagrams for Viola & Jones and CAMSHIFT algorithm respectively.

III. Experimental Results

Solution proposed by this paper gives an algorithm which can detect face(s) in the video in real time. Figure 4 and figure 5 show the output of single and multiple face detection.
Video of face detection can be found on youtube[13].

Fig. 4. single face detection

This algorithm has been tested on video given to us as well as Video 6. It works well on both of them. It has

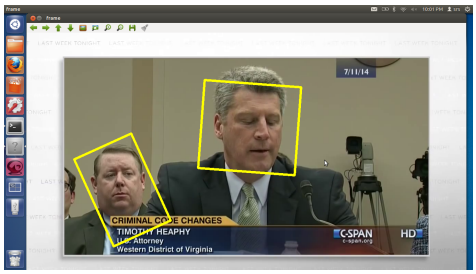

Fig. 5. multiple face detection

also been tested on Benchmark video.The limitations on Benchmark video is stated in the subsequent section.

A. Limitations

Though the algorithm is working with high speed face detection and it gives very efficient output, our algorithm is not capable in the scenario where background and face colour difference is very low. In such cases, the algorithm might give larger box over the face, it might not detect the face at all or it may detect a non face image.

The reason of above phenomena is- because of adaptive CAMSHIFT, i.e, if mean of background and foreground is more or less same, then it will increase the ROI to include that background also considering it as a skin colour. It is also possible that, if after detecting the face, in one of the faces, if the mean gets more similar to the ROI in previous frame then the algorithm will force the box to reside over a region which might not be a face at all.

Figure 6 and figure 7 show this misidentification of face within the frame with less background and foreground colour difference.

The proposed algorithm fails on Benchmark video, because of the incapability of Vila & Jones to detect faces with props like glasses, cap, beard etc. It fails most of the time in Benchmark video to detect a face in a frame. Hence if face is not detected, the CAMSHIFT will never get called and tracking will never happen.

Fig. 6. unidentified larger box because of low background and foreground colour difference

Fig. 7. non face identification because of low background and foreground difference

IV. PERFORMANCE ANALYSIS

The algorithm is implemented in python using OpenCV libraries. It has been implented over a machine with specifications of 4 core processor and 4 gb ram. The resulting output video gives us a speed of more than 40($precisely 42$) fps with the same resolution as the input video which suffices the goal of face detection in real time.

A. Time Analysis

The time complexity of Viola and Jones Algorithm used for face detection is $O(5n^2)$ (where n is the dimension of the input image)[12] and that of CAMSHIFT algorithm used for face tracking is $O(\alpha n^2)$ (where n is the size of ROI window and alpha= some constant) [8]. Hence the overall complexity of our algorithm would be much less than the Viola and Jones algorithm as it is called for only 1/29 frames of the total frames of the video.

Worst case for this algorithm will be a video in which there is no face because in such situation subsequent frames are going to go through Viola and Jones algorithm only in order to find face(s) in an image. This will degrade the performance of the algorithm because of high computational complexity of Viola & Jones algorithm.

So proposed algorithm incorporates the advantages of Viola & Jones as well as CAMSHIFT in beneficial manner.

V. CONCLUSION AND FUTURE WORK

This algorithm fulfills the goal of the project which was identifying faces in 1280*720 video with 30 fps in real time. Though in the case of no face in subsequent frames, detection gets slow in time. We need to solve this problem.

One of the possible solution is to do parallel processing in subsequent frames in condition of frames detected. Multiprocessing would depend on the number of cores in the CPU. For example, let us assume a CPU has 4 cores. The task of detecting the face in a frame could be distributed to the 4 cores independently and the results could be combined. Another approach through multiprocessing would be to give input of 4 frames to 4 different cores and detect a face in each of them. The former approach would seem to be more feasible as we are detecting face for just 1 frame (rather than 4 frames), and once the frame is detected- 29 frames are used to track it.

VI. ACKNOWLEDGMENT

We would like to acknowledge Prof. Mehul Raval, Prof. Ratnik Gandhi and Mr. Dhruv Gupta for their constant guidance and valuable inputs.

REFERENCES

[1] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57.2 (2004): 137-154.
[2] Jog, Aditi, and Shirish Halbe. "Multiple Objects Tracking using CAMshift Algorithm in Open CV."
[3] Yang, Ming-Hsuan, David Kriegman, and Narendra Ahuja. "Detecting faces in images: A survey." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24.1 (2002): 34-58.
[4] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.
[5] http://www.computervisiononline.com/blog/tutorial-using-camshift-track-objects-video
[6] http://opencv-python-tutroals.readthedocs.org/en/latest/py_tutorials/py_video/py_meanshift/py_meanshift.html
[7] http://www.mukimuki.fr/flashblog/2009/06/18/camshift-going-to-the-source/
[8] Emami, Ebrahim, and Mahmood Fathy. "Object tracking using improved CamShift algorithm combined with motion segmentation." Machine Vision and Image Processing (MVIP), 2011 7th Iranian. IEEE, 2011.
[9] Mohammed, Abdulmalik Danlami, and Tim Morris. "An Improved CAMShift Algorithm for Object Detection and Extraction."
[10] http://www.ricoh.com/technology/rd/column/020.html
[11] Pissarenko, Dimitri. "Eigenface-based facial recognition." December 1st (2002).
[12] https://sites.google.com/site/5kk73gpu2012/assignment/viola-jones-face-detection
[13] https://www.youtube.com/watch?v=lY6zYZ5-Zd8