

Masked Face Detection using the Viola Jones Algorithm: A Progressive Approach for less time consumption

<https://doi.org/10.3991/ijes.v6i4.9317>

Aishwarya Radhakrishnan Nair ^(✉), Dr. Amol D. Potgantwar
Savitabai Phule Pune University, Pune, India
arn121294@gmail.com

Abstract—The use of CCTV surveillance is today's need in public and private sector for ensuring security against terrorism and robbery. Regular expressions are used to signify enormous sets of motion attributes captured in video. The video vigilance is popular system without using human interference to capture important scenes. The motive of the work is to introduce auto-matic revelation of masked objects in real time with a surveillance camera. The main aim is to detect masked person automatically in less time period. In this paper, the researcher proposes a system that consists methods which uses four variant steps that are the steps of calculating distance range of person from the camera, eye or vision line detection and face part detection such as mouth detection and face detection. Performance of proposed algorithm is carried out on various real time inputs. Experimental evaluation shows that proposed algorithm exceeds better in terms of time consumption. This unique approach for the problem has created a method transparent and easier in complexity so that the real time implementation can be made beneficial and workable. Analysis of the algorithms fulfillment on the test video track gives appropriate judgments for additional improvements in the masked face detection performance. Finally, based on the research, the axioms were useful for the work which can be usually accessible from available algorithms.

Keywords—CCTV, Face Detection, Masked face Detection, Eye detection, Video analytics

1 Introduction

Sometimes CCTV vigilance system is required in today's insecure world where there is robbery in banks, homes and other important places and also terrorist attacks are happening in open and private areas. There are some security issues faced while identifying or detecting any suspicious person if they wear mask. Self-analysis of video is active to recognize and successfully grab suspect person. However, while it is relatively easier to monitor, but for this monitoring, expensive system is required. It is difficult and also very much time consuming to examine the video recordings or feeds by using protection guards. Currently CCTV surveillance method details are located

for while not tracking in order to take advantage of the video recording on a forensic.[1]

Several challenges may face from the occurrence of video surveillance systems. They are as follows:

- Illumination changes (brightness): A day light scenario contains brightness that are continually change and that affects on video recording. This can occur during day hence due to this video quality compromises with illumination.
- Live background: A natural scene generally comprises of dynamic objects. These dynamic contents can be made up by different trees, changing surface waters, waving flags, etc.
- Moving object: It can prevent the object in a scene that is moving when it is captured in video due to this it creates ghost illusion. For instance: Suppose a parked vehicle may be a car leaving the scene, the equivalent sector should be accepted as a segment of the frame or background.
- Noise in the Video: The images in the video may consist of brightness fluctuations or color fluctuations in video sequences, which are called noises. Sometimes video contain different types of noises, such as sensor noise or compression artifacts that affect the video quality.[2]

CCTV system is primarily used to monitor public areas such as bank, shopping complexes and bus stop and in private areas such as corporate sectors. Old security system preferences human to evaluate CCTV screens. At the same time far more advance systems are likely to spot and warn all the controller on situations that create probable security measures pitfalls, they have limited functionalities. [3]

A good elevated dark channel most recent device included by means of Gaussian chart can be proposed. In this an inverted reduced mild image is believed just like a blur image and removing technique which is not really a excellent enhancement one. Night video advancement practices convey the mobile subject and transform to avoid noisy part. A delighting approach uses several day time and evening time images, but the increased productivity is not clear. A development technique for complicated illumination condition finds the source of damage and boost the video Low light video development methods [6] .

General mask detection method handles complicated algorithms like function centered methods and learning centered algorithms. Approaches based on highlights of encounter manipulate the in-formation involving skin qualities like skin tone to decide whether there may be obstruction on face or not. The propose SVM for curb face recognition. The approach for this faces detect with gabor wavelets, PCA. Having said that, these approaches are generally demanding. Here we assess a few of the measures of implementation of criminal face detection[4].

This paper is formulated as : First there is brief introduction. In second section survey of related work is discussed. While the third section introduces the system architecture. System Analysis is done in section four which includes Algorithm and mathematical model. Performance analysis is explained in the fifth section. Conclusion of the paper is framed in section six.

2 Related work

In the scene whether a person is present anywhere in the front of camera or not is main question so for this the area width between person and camera needs to be calculated in the form of increasing or decreasing distance. A pinhole digicam prototype is used. Pinhole digital camera prototype or version is extraordinarily simple wherein, light enters originating from the scene or distant items, only just one beam intrudes from any distinct specific point. This particular point is later flashed on imaging surface. Eye axis or line feature is diagnosed on productivity window. Eye-brows will be areas having low level in comparison with some other elements of face; their spots can match to any local points of the horizontal projection histogram. Therefore the eye axis or line detection algo-rithm can be squeezed as dictate procedure on the outside gray value projection histogram. Facial part detection is achieved in two parts. Face detection and also diagnosis associated with cosmetic locations like eyes, nasal area and mouth is fond out by algorithm designed by Viola Jones. Face portion detection normally takes more hours to execute. [4]

A significant system is designed for identify activities in real time from video streams automatically. Regular expressions are applied to this system to serve unlimited sets of motion attributes from a video. It constantly handles trajectory-based along with regularly articulated activities and also for faster recognition, it grants algorithms of polynomial time graph. The regular expressions representing motion properties either be provided automatically from non-negative examples of strings using offline automated learning method. Confidence calculations are related with recognition that uses Levenshtein distance between a string representing a motion signature and the regular expression describing an activity. [5]

The identification is done in real time. We have proposed two approaches: The first approach points on two basic steps: creating the background to set the static image of the scenario and difference the background from the front image, which allows getting moving objects. The second approach apply directly on captured images. The step of edge detection is utilized to capture the edges of objects only. Every approach has some limitations. [2]

A system for surveillance that also consists of video analysis rele-vant social media information. The work is still in the under progress but will developed the background IP such as the video search system which was been tested on the TRECVID 2010 standard video dataset and the system gain the best ranking under the task of video known-item search as discussed above, as well as the face identity inference algorithm which is secure, safe, effective and efficient for surveillance environment.[1]

Closed-circuit television (CCTV) also known as video surveillance is the atmost useful technology that is used maximum in the field of security purposes. CCTVs are nowadays observed at many places ranging from public places to various private places. One of the most crucial and challenging issue in installing the CCTV cameras on a large scale is space that occupied for storage of the footage. Footages are mostly stored in the secondary storage devices such as hard disk drives. So, to reduce the storage space, compression techniques are applied. An algorithm is design for storage

optimization especially for CCTV i.e. Closed-Circuit Television as storage is a real challenge with increasing market demand for the third eye at almost every specific place in the city.[6]

Due to the variety of the solutions the following categories have been taken into consideration: systems based on object detection, tracking and movement analysis, systems able to warn against, detect and identify abnormal and alarming situations, systems based on vehicle detection and traffic or parking lots analysis, object counting systems, systems based on multiple integrated camera views, privacy preserving systems and systems based on cloud environment. The paper describes several solutions for each category and underlines main functionalities of the current intelligent surveillance systems. [7]

General mask detection or scarf detection algorithms deal with complex algorithms like feature based algorithms and learning based algorithms [9]. Methods based on facial features exploit the information of facial features such as mouth [10] or skin color [11] to decide whether there is occlusion on face or not. Jia H. and Martinez A. M. propose support vector machines for occluded face recognition. Min. R. [13] approach occluded face detection with gabor wavelets, principle components analysis and support vector machines. However, these techniques are computationally intensive. Here comparison of some of the steps of implementation of masked face detection is done.

In the first step of masked face detection, eye line is detected. The eye line detection algorithm can be reduced as a valley finding procedure on the horizontal gray value projection histogram [11]. As video analytics deals with detection of person and events like walking, falling etc., we make use of the fact that person and face detector are present in the system. We consider person detector implemented by N. Dalal and B. Triggss method of Histogram of Oriented Gradients [14].

The Viola Jones face detection algorithm has four stages, namely, Haar Feature Selection, Integral Image creation, Adaboost Training and Cascading Classifiers [18]. Viola Jones face detection procedure classifies images based on the value of simple features. There are three features, namely two rectangle, three rectangle and four rectangle. These rectangle features can be computed very rapidly using an intermediate representation for the image which is called as the integral image.[19] Analog Devices Inc.'s Cross Core Embedded Studio (CCES) can be used along with HOGSVM for person detection and distance of person from camera step[20] [21].

3 System architecture/system overview

The use of CCTV surveillance is today's need in public and private sector for ensuring security against terrorism and robbery. Regular expressions are used to signify enormous sets of motion attributes captured in video. The video vigilance is a popular system without using human interference to capture important scenes. The motive of the work is to introduce automatic revelation of masked objects in real time with a surveillance camera. The main aim is to detect masked person automatically. This system consists of methods, which use four variant steps of calculating distance

range of person from camera, eye or vision line and face part recognition and eye or vision detection. The axioms were useful for that work which can be usually accessible from available algorithms.

4 System analysis

4.1 A Proposed System

Functions like walking, falling etc. of individuals are noticed with the help of video analytics, we utilize the fact that person and face detector are present in the system. We acknowledge person detector carried out by Histogram of Oriented Gradients. Set of features based on working out well-normalized regional histograms involving image gradient orientations in a compressed grid is known as Histogram of Gradients (HOG). Satisfying results for person recognition, reducing fake positive rates in accordance with the most effective Haar wavelet based detector is obtained. Detecting whether an individual is wearing mask or not with a mask on the face is the main target.

The four steps to be taken into consideration are:

i) Distance of person from Camera: The suitable approach to acknowledge if person is approaching in the direction of your camera or even going away is to locate out the yardage in between those along with the camera. Because person is getting close to the camera, distance concerning particular person and also video camera will probably lower and also deal with discovery is usually provoked. Pinhole camera model is utilized to discover the distance among particular person and also camera. Pinhole video camera unit is easy video camera unit through which, ray of light comes in from the scene or far away gadgets, but only a one single ray enters from any specific point. This point is then visualized onto an imaging surface.

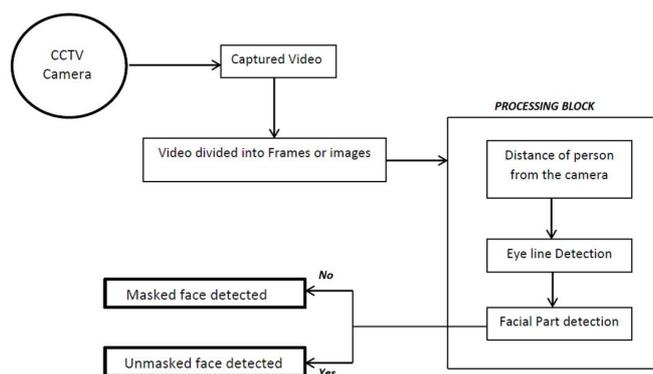


Fig. 1. System that detects mask on a person's face using Viola Jones Algorithm

ii) Eye Line Detection: In this kind of stage of criminal face detection, the eye line of a person is recognized in resultant screen of individual detection. Eyebrows along

with the eyes of a person are portions with minimal gray-level as compared to other areas of the facial skin; their positions must correlate with the regional valley of the horizontal projection chart. Therefore the eyes series detection criteria is often miniaturized being a area finding method about the horizontally gray value projection histogram.

iii) Facial Part detection: Recognition of parts of face based covered face recognition is accomplished in two parts. Face recognition accompanied by facial part detection. Face recognition furthermore recognition of facial parts such as eyes, nasal area and mouth is attained by algorithm designed by Viola Jones.

iv) Face Detection: Cosmetic component detection takes in a longer period to complete the task (600 seconds to deal with 525 frames). Therefore we propose this task of face detection, which takes somewhat less time (360 seconds to deal with 525 frames). This is accompanied by application of algorithm which deals with face detection. If eyes are recognized and later if face is recognized, it signifies that there's no disguise on the person's face. If eyes are recognized but face is not recognized, it signifies that individual has put a cover on rest of the face.

Algorithm used in the proposed System:

Viola Jones Algorithm: The Viola Jones object detection framework is the object detection framework to provide competitive object detection rates in real-time proposed by Paul Viola and Michael Jones. Although it can be trained to detect a variety of object classes, it was motivated primarily by the problem of face detection.

- 1) Given examples images $(x_1, y_1), \dots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples.
- 2) Initialize weights $w_{1,i} = \frac{1}{21m} ; \frac{1}{21l}$ for $y_i = 0, 1$, where m and l are the numbers of positive and negative examples
- 3) For $t = 1, \dots, T$:
- 4) Select the best weak classifier with respect to the weighted error:
- 5) $t = \min_f; p; \sum_i w_{i,j} h(x_i; f; p;) y_i$
- 7) Define $h_t(x) = h(x; f_t; p_t; t)$ where f_t, p_t and t are the minimizers of t
- 8) Update the weights:
- 9) $w_{t+1,i} = w_{t,i} \frac{1 - e_i}{e_i}$
- 10) where $e_i = 0$ if example x_i is classified correctly and

$$e_i = 1 \text{ otherwise, and } t = \frac{1}{t}$$

- 11) The final strong classifier is:

$$C(x) = \sum_{t=1}^T w_t h_t(x) - \frac{1}{2} \sum_{t=1}^T w_t$$
 where $w_t = \log \frac{1}{e_t}$

4.2 Mathematical Model

Input: Camera as a input for face detection Output: Masked face detection in video

S = It holds the system parameter.

Let S be a system,

Such that $S = \{P, C, E, F, D, I\}$ Where, P represents set of persons in video:

$P = \{p_0, p_1, p_2, p_3, \dots, p_n\}$

C represents distance from camera:

$C = \{c_0, c_1, c_2, c_3, \dots, c_n\}$

E represents eye line detection:

$E = \{e_0, e_1, e_2, e_3, \dots, e_n\}$

F represents facial part detection:

$F = \{f_0, f_1, f_2, f_3, \dots, f_n\}$

D represents detection of eye lines:

$D = \{d_0, d_1, d_2, d_3, \dots, d_n\}$

I represents face detection/undetected:

$I = \{i_0, i_1, i_2, i_3, \dots, i_n\}$

O = Set of detected masked cover faces

5 Performance analysis

5.1 Dataset

As the video is captured by the CCTV cameras or the web cameras, no dataset is required and used in this implementation. The processing is done over Real-time videos. The real-time videos that are captured are divided into smaller frames and the processing is done.

5.2 Implementation

The proposed system demonstrates the process of detection of people's face to identify whether it is masked or not. This procedure undergoes various steps such as distance calculation, detecting the eyeline, detecting the face and finally detecting the facial part. As compared to the existing system, the time that these processes require to complete is less in the proposed system. This allows us to analyze the results faster and easier.

Result: Method of face recognition designed by Viola Jones categorizes images on the basis of the worthiness of easy attributes and features. There are around three attributes, specifically two rectangles, three rectangles and four rectangles. A two-rectangle function value is calculated or enumerated by involving the supplement of

the pixels within two rectangle regions. These parts must be exclusively of related measurement and form and are also horizontally or vertically alongside each other. Moreover, options that come with three and four rectangle may be computed. These rectangle characteristics might be calculated quickly utilizing a transitional rendering for the picture, which will be called the built-in image or integral image.

Range of the person from camera approach is applied to recapture or examine if person is approaching towards the camera or going far from the camera. Eye range recognition shows valley in horizontal histogram projection. If vision range is found, face recognition approach may be properly used to see if person has worn a mask on his face or not. Artificial catching rate is the highest level of during total eye path catching criteria in addition to soon after during total eye detection. This really is due the fact that vision recognition and vision point recognition discover modest section in picture and for photographs with bad or reduced solution; that recognition won't be appropriate or specific resulting in fake detections. Performance time of face portion recognition is probably the most when compared with most of the different outstanding measures because it grips face recognition used by recognition of the areas of the facial skin which is really a hard algorithm. To study the efficiency of the stated four measures, two films are tried; with disguise on face and without disguise on face and precision was computed as (Number of Correct detections/Expected detection).

The real time result analysis graphs for face detection rate and the time comparison is shown in the figures given below. In the graph that shows Detection rate comparison 'X' axis contains face detection rate and 'Y' axis contains the performance of existing system and the performance of proposed system.

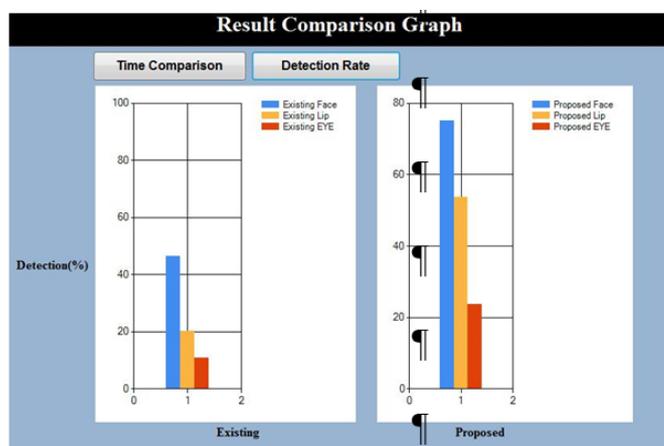


Fig. 2. Performance analysis of existing algorithms and proposed algorithm that compares the detection rate of masks

Similarly, In the graph that shows Time comparison 'X' axis contains the time required for face detection and 'Y' axis contains the performance of existing system and the performance of proposed system.

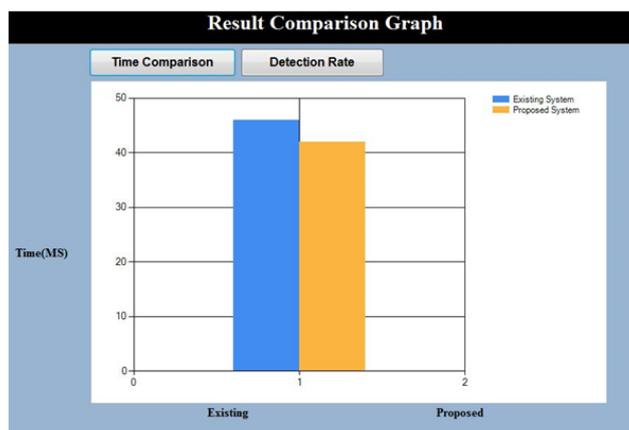


Fig. 3. Analysis and comparison of Time required for detection of mask between existing algorithms and proposed algorithm

6 Conclusion

Masked face detection goes from different stages and analyzed. As compared to others calculating the distance of person from camera is more robust and correct. Eye line detectors now is easier for us to implement nevertheless however it contributes to detections that are fake within inadequate quality images. Eye feature detection is reliable for identifying eyes on face. Facial component discovery can be sturdy along with time-consuming step due to many regions. This planned process in phrase with accuracy, reliability and less of time consumption along with tenderness with uncovering shows that particular planned algorithm accomplishes superior result.

7 Acknowledgment

I would sincerely like to thank our Head of Department Prof. Dr. Amol Potgantwar, Department of Computer Engineering, SITRC. Nashik for his guidance, encouragement and the interest shown in this project by timely suggestions in this work. His specialist recommendations and scholarly feedback had considerably increased the potency of this work.

8 References

- [1] Lekha Chaisorn, Yongkang Wong—Video Analytics for Surveillance Camera Networks; IEEE 2013
- [2] Hanane Belhani, Larbi Guezouli,—Automatic detection of moving objects in video surveillance; 2016 Global Summit on Computer & Information Technology..

- [3] Junn Min Pang, Vooi Voon Yap, Chit Siang Soh ,—Human Behavioral Analytics System for Video Surveillance; 2014 IEEE International Conference on Control System, Computing and Engineering, 28 - 30 November 2014, Penang, Malaysia.
- [4] Gayatri Deore, Ramakrishna Bodhula, Dr. Vishwas Udpikar, Prof. Vidya More, – Study of Masked Face Detection Approach in Video Analytics; Conference on Advances in Signal Processing (CASP) Cummins College of Engineering for Women, Pune. Jun 9-11, 2016.
- [5] M. Karki, S. Basu, R. DiBiano, S. Mukhopadhyay, J. Weltman, M. Stagg, — A Symbolic Framework for Recognizing activities in Full Motion Surveillance Videos; IEEE 2016. <https://doi.org/10.1109/SSCI.2016.7850118>
- [6] S. Arora, K. Bhatia, Amit V,—Storage Optimization of Video Surveillance from CCTV Camera; 2016 2nd International Conference on Next Generation Computing Technologies (NGCT-2016).
- [7] H Liu, S. Chen and N. Kubota,—Intelligent Video Systems and Analytics: A Survey; IEEE Transaction on Industrial Informatics, Vol. 9, No. 3, pp. 1222-1232, Aug. 2013. <https://doi.org/10.1109/TII.2013.2255616>
- [8] Rui Min, Angela D'Angelo, Jean-Luc Dugelay,— Efficient scarf detection prior to face recognition; EUSIPCO, 18th European Signal Processing Conference, Denmark, France, pp. 259-263, Aug. 2010.
- [9] Wright J., Yang A. Y., Ganesh A, Sastry S. S., Ma Y.,—Robust face recognition via sparse representation; IEEE Transactions on pattern Analysis and Machine Intelligence, Feb. 2009, Vol. 31, pp. 210-227. <https://doi.org/10.1109/TPAMI.2008.79>
- [10] C.Y. Wen, S.H. Chiu, Y.R. Tseng and C.P. Lu,—The Mask Detection Technology for Occluded Face Analysis in the Surveillance System; Journal of forensic sciences, Vol. 50, no. 3, pp. 593-601, 2005. <https://doi.org/10.1520/JFS2004409>
- [11] D. T. Lin, M. J. Liu,—Face Occlusion Detection for Automated Teller Machine Surveillance; Lecture Notes in Computer Science, Vol. 4319, 641-651, Sept. 2006. https://doi.org/10.1007/11949534_64
- [12] Jia H. , Martinez A. M.,—Support vector machines in face recognition with occlusions; Proceedings of the IEEE Computer Society conference on Computer Vision and Pattern Recognition Workshops, June 2009, 136-141. <https://doi.org/10.1109/CVPR.2009.5206862>
- [13] Min. R, Hadid A, Dugelay J. L.,—Improving the recognition of faces occluded by facial accessories; Proceedings of the IEEE International conference on Automatic Face and Gesture Recognition and Workshops, March 2011, pp. 442-447.
- [14] N. Dalal, B. Triggs,—Histograms of Oriented Gradients for Human Detection; IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, Vol. 1, pp. 886-893. <https://doi.org/10.1109/CVPR.2005.177>
- [15] G. Bradski and A. Kaehler,—Learning OpenCV; O'Reilly Media, Inc., 2008, pp. 370-404.
- [16] Blackfin,—HOG SVM Detector Product Reference Guide; Analog Devices, Inc., 2015
- [17] Min-Quan Jing and Ling-Hwei Chen,— A novel method for horizontal eye line detection under various environments; Int. Journal of Pattern Recognition and Artificial Intelligence, Vol. 24, No. 3, pp. 475-498, 2010. <https://doi.org/10.1142/S0218001410008020>
- [18] P. Viola and M. Jones,—Robust Real-Time Face Detection; Int. Journal of Computer Vision, Vol. 57, no. 2, pp. 137-154, May 2004. <https://doi.org/10.1023/B:VISI.0000013087.49260.fb>
- [19] Viola, Paul and Michael J. Jones,—Rapid Object Detection using a Boosted Cascade of Simple Features; Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, Vol: 1, pp. 511-518. <https://doi.org/10.1109/CVPR.2001.990517>

[20] J. Fernandez, S. Kottekkode,—Object Detection; U.S. Patent 13/888,993, May 15, 2014.

[21] A. Sripadarao, B. Poyil,—Facial Detection; U.S. Patent 14/013,122, Nov 3, 2015.

9 Author

Aishwarya Radhakrishnan Nair is a PG scholar at the Department of Computer Engineering at SIRTIC, Savitribai Phule Pune University, Pune India.

Dr. Amol Potgantwar is the Head of Department of Computer Engineering at SIRTIC, Savitribai Phule Pune University, Pune India.

Article submitted 31 July 2018. Resubmitted 17 August 2018. Final acceptance 29 August 2018. Final version published as submitted by the author.