Implementation of Viola-Jones for Detection of Facial Factors of Human for Prospect of Image Recognition

Abdul Wasay Siddiqui[1], Prof. M. Afshar Alam[2], Dr. Harleen Kaur[3], Dr. Jawed Ahmed[4]
{wasay.siddiqui8@gmail.com[1], aalam@jamiahamdard.ac.in[2],
harleen.unu@gmail.com[3], jahmed2047@jamiahamdard.ac.in[4]}

Bachelor of Technology, Computer Science and Engineering,
Jamia Hamdard University, New Delhi 110025, India[1].
Department of Computer Science and Engineering,
Jamia Hamdard, New Delhi 110025, India[2,3,4].

Abstract - This is an application for tracking and detecting human faces and eyes from video or images, which can be used for different purposes and applications. The face is one of the distinguishing factors which can easily be used for the identification of an individual. The primary goal of this paper is to study the Viola-jones (VJ) algorithm, its four stage processes (Haar like Features, Integral Images, Adaboost and Cascading) and its implementation using OpenCV for face, eye and smile detection. It's an adroit algorithm basically targets the frontal faces for better detection accuracy.

Keywords: Face Detection, Viola- Jones, Haar Features, AdaBoost, Cascading, Integral Images, Rekognition, Deep-Face, Face-Net, Machine Learning, Deep Learning, Computer Vision.

1 Introduction

In the past few decades, significant work has been done in the field of machine learning and deep learning, which resulted in the most promising and demanding application in the field of computer vision, **Face Detection**. Face detection has a robustness to focus computational resources of face images and can be considered a substantial part of face recognition operations. Sometimes face detection could be complicated and intense with pictures due to the external varying factors such as differences in lighting conditions, human poses, expressions, image resolution, the presence of spectacles, positional orientations, facial hair, skin colours and capturing angle. Lots of study and proposed work in the very same field made face detection quite advanced and accurate. After the introduction of **Viola-Jones (VJ)** algorithm real time detection is possible with ease. We have done detection using VJ as its one of the most powerful and robust algorithms at its simplicity and being used for computer vision since 2001. It's famous not only for detecting faces in images but also for real time computer vision face detection in videos or from webcam. Viola-Jones (VJ) algorithm comprise four stages: **Haar like Features, Integral Images, Ada Boosting** and **Cascading**. Therefore, with the use of viola jones algorithm plays a crucial role of detection in many applications such as **object detection, face recognition system, surveillance of videos** and many more. Face detection

technology has proliferated and is currently being used to make the world faster, smarter, safer and more convenient.

2. RELATED WORK

Huge organisations also play a dominant role in the mix. Most of the gigantic software organisations and web companies are proposing their discoveries, implementing and analysing them in the fields of artificial intelligence, machine and deep learning, image recognition, face detection in an attempt to further understand as rapidly as possible. **DeepFace,** is a program which is used to check whether faces in two photographs are of the same person or not, with an accuracy of 97.25%, launched by worldwide known organisation Facebook. Humans answered correctly with accuracy of 97.53% in the same test. Or just 0.28% better [1]. Google went one better with **FaceNet**, attain a record-breaking **accuracy of 99.63%** (0.9963 ± 0.0009). **FaceNet** technology is very famous and implemented in Google Photos where its task is to sort images and tag them automatically on the basis of recognition of the people in the pictures. It has played an important role in the biometrics landscape and was quickly followed by the online release known as OpenFace [2]. Zhong-Qiu Zhao, Peng Zheng and other IEE fellows proposed a deep learning approach for visually detecting object which is capable of processing images rapidly and achieving high detection rates [3]. Faizan Ahmad, Aaima Najam and Zeeshan Ahmed evaluated various face detection and recognition methods, provide solution for image facial detection and recognition with higher accuracy [4]. Amazon provided **Rekognition** which is a cloud-based service for face recognition to agencies of law enforcement [5]. This approach is used to recognize hundreds of people in an image and can also perform face matching against databases of millions of faces. Ensemble of algorithms is used as a feature-based approach to face detection. Edge features and its related chrominance values are used to differentiate the image as skin and non-skin regions. Shylaja S S, K N Balasubramanya Murthy showed the use of Laplacian of Gaussian (LoG) filter as an edge detector which is quite sensitive and appropriate when images are mostly full of noise and have multiple faces in them [6].

3. PROPOSED WORK

We will be understanding the working and implementation of the Viola-Jones on the basis of its four staged processes.

3.1. Viola Jones

Working of VJ algorithm was proposed by Paul Viola and Michael Jones [7], mostly used for real time face detection in image or video. It produces the best result on frontal faces and was mainly designed for it. It converts the colour image into grayscale image (less data to process) and scans the whole image to find basic features like eyebrows, eyes, nose, cheeks, mouth chin and many more by iterating a box all over the image. Face will be detected by bounding a box around it when all these features are found and captured in that bounding box at the same instance. And it'll do the same and find the features at exactly the same positions in the colour image.

Viola Jones consists of four stages:

Haar like Features

Its digital features used to detect human face features [8].

Integral Images

It's a matrix used for faster computation of Haar like Features.

Ada-boost

It's a process where features (weak classifier) are combined to produce a strong classifier for detection.

Cascading

It's a hack to speed up the process of Adaptive Boosting.

3.2. Libraries Used

3.2.1 OpenCV

OpenCV is a deep learning library. It contains optimized algorithms more than 2.5k [9]. Detection and recognition of faces, identification of objects, classification of human actions in videos, tracking movement of the objects, extracting 3-Dimensional models of objects, producing 3-Dimensional point clouds from stereo cameras, clubbing images to produce image of an entire scene with better resolution quality, searches image database for detecting similar images, removing red eyes, capturing movements of eyes, establishing markers to overlay it with augmented reality can be achieved using the algorithm available in OpenCV. It's used to build bounding boxes around the detected faces, eyes and smiles as shown in fig 1.
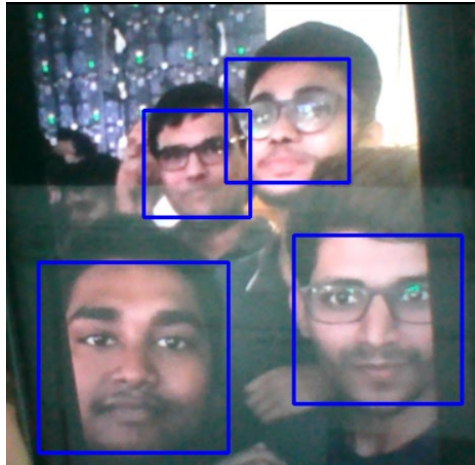

Fig1: Bounding boxes around faces

4. IMPLEMENTATION

Detection process has been implemented using Haar like Features, Integral Images, Adaboost and Cascading.

4.1. Haar like Features

The famous Hungarian mathematician of the 19th century, who developed mathematical concept Haar Wavelet. Object recognition uses these as digital image features. These are derived from Haar Wavelets and that's why they are called Haar like features, they are kind of descendant. These features are mainly classified into three features: Edge Features, Line Features and Four Rectangle Features.
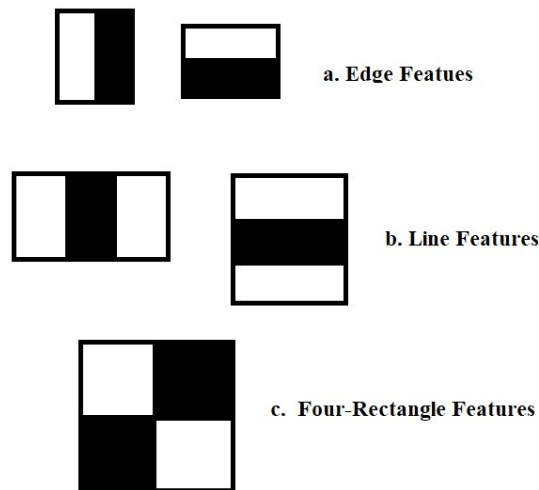
Fig2: Haar-like features

Haar like Features was proposed by Viola-Jones to detect facial factors. This is used for finding the location of the human faces in an image. There are some common and universal properties in all human faces which can easily be captured from naked eyes as the lips central region of lips are darker than their neighbouring pixel values and eyes region is lighter than the nose.
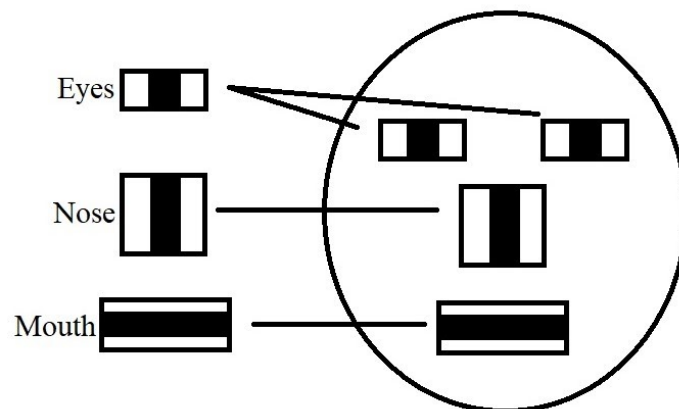
Fig3: Working of Haar like Features on face model

The haar-like-feature algorithm is also used for feature selection and extraction of an object in images, with the help of edge detection, line detection, centre detection, four rectangle detection for detecting eyes, nose, lips cheeks, etc. in the picture. It selects the important features in an image and extract these for detecting faces. Haar Like Features basically detecting human face features on the basis of facial factors as shown in Fig3. These features are matrices of pixel with values between 0 to 1 (example 0.5 or 0.12) on the basis of colour intensity. Where white is almost close to 0 and black is almost close to 1.

4.2. Integral Images

Viola Jones has used the concept of summed-area table which is an algorithm of generating efficiently summed values of the grid [10]. Also known as Integral Images in the field of image analysis and processing. At its simplicity it's a multidimensional matrix of the same size as the original image. Where the values of integral images are calculated by adding up all the top-left region values of the original images (pixel values) as shown in fig 4 and fig 5.

1	2	5	7	2
9	8	0	4	9
7	6	10	2	0
3	8	1	5	4
9	5	0	1	3

Fig4: Original Image

1	3	8	15	17
10	20	25	36	47
17	33	48	61	72
20	44	60	78	93
23	58	74	93	111

Fig5: Integral Image(I)

This eases up and simplifies computing the sum of rectangular areas in an image to calculate the value of haar like features at that area by following the formula as shown in equation (1).

$$sum= I(a)+I(c)-I(b)-I(d) \tag{1}$$

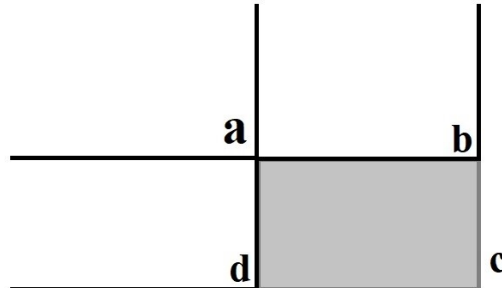In Fig 6 a, b, c, d belong to the integral image I.

Fig6: Sum of area enclosed by abcd.

4.3. AdaBoost

It's a process of summing up weak classifiers(features) to make a one robust Classifier. We are using adaboost to avoid 180,000+ features detection, because each feature (line or edge or four rectangle) is trying out all possible and different positions to detect in an image.

$$F(X) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \alpha_4 f_4(x) + ... \qquad (2)$$

In the equation (2), α is weight corresponding to the features (eyes, nose, lips, etc) f(x), also known as weak classifier in individual. Where F(X) is a Strong Classifier. This is a strategic selection approach of features, where $f_{i+1}(x)$ feature(nose) complements preceding features(eye) $f_i(x)$ to generate a strong classifier for proper detection with less errors [11].

4.4. Cascading

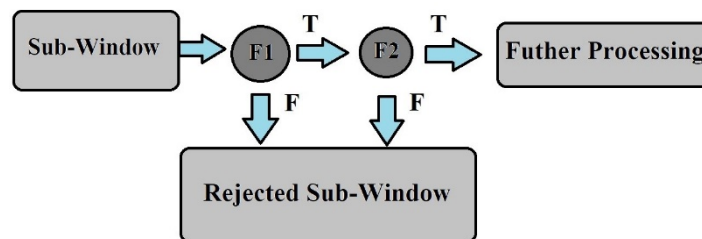It's a process to speed up the detection training of images.

Fig7: Series of classifiers.   .

Fig7 F1 and F2 are the features like nose, eyes, lips etc. to detect the human face. If F1 is the feature that is required to detect the human frontal face then only flow of control will move to F2, otherwise it will go to the Rejected Sub-Window. This hack will optimize the process of detection and decrease the computation as only required features will be considered and others will not, like hair, neck etc.

5. TRAINING CLASSIFIER

For training the classifier we have follow sequentially all the four stages of Viola Jones algorithm. In the process of training the image is shrunk and scaled down to 24px X 24px to apply Haar Features to the images. Whereas in real time we do not scale down the images, wherever we scale the features up. So those features that are important we scale them up and look for them on the images [12]. To train we have to provide data with lots of images of the frontal face of different people and also have to provide some non-face images. We have used pre-trained Haar Cascade Classifiers for Frontal face detection, eye detection and smile detection.

6. RESULTS

After implementing all the four staged processes of Viola-Jones algorithm and developing this application we tested detection in real time as well as on captured images or photos through mobile camera. And we got highly satisfying results with high accuracy. Where most of the facial factors have been detected. As we can estimate from Fig 8 that maximum number of prominent faces are easily detected in a group image.
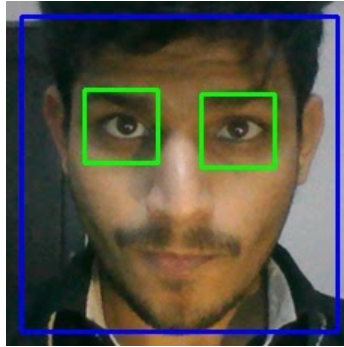

Fig8: Faces are detected in an image   .
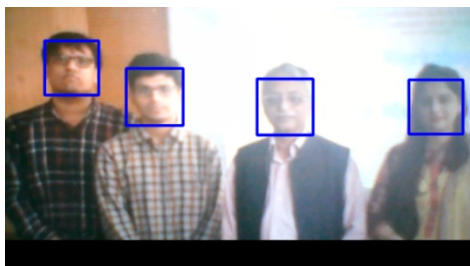
Fig 9: All the faces have been detected in blurred image

From Fig 9, we can deduce the highly satisfying accuracy of Viola- Jones as it detects all the faces in blurred images and bound them in blue colour bounding boxes.

Fig10: Face and eyes Detection from webcam.

In Fig 10, we've shown the real time face and eyes detection using a web camera where face and eyes are bounded by blue and green bounding boxes respectively.

7. CONCLUSION

We have presented an approach of face detection using Viola-Jones and pre trained classifier with high detection accuracy and low computation time. Using this approach, we can detect multiple faces, eyes and smiles in images and videos. Most importantly it can be used for real time detection. Due to real time detection this application can be quite helpful in different domains like Traffic Control, Banks, Hospitals, Residential and many other areas for security and safety purposes.

REFERENCES

[1] Yavin Taigman, Ming Yang, Marc'Aurelio Ranzato and Lior Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," Conference on Computer Vision and Pattern Recognition (CVPR).

[2] Florian Schroff, Dmitry Kalenichenko and James Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015.

[3] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu, Fellow, IEEE, "Object Detection with Deep Learning," IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS FOR PUBLICATION

[4] Faizan Ahmad, Aaima Najam and Zeeshan Ahmed, "Image-based Face Detection and Recognition".

[5] Amazon Rekognition, "https://aws.amazon.com/rekognition/,".

[6] Shylaja S, K N Balasubramanya Murthy, S Natarajan Ramya S, Radhika S, Nikita J and Prateek R, "Face Detection in Color Images Using Skin Color, Laplacian of Gaussian and Euler Number."

[7] Paul Viola and Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Conference on Computer Vision and Pattern Recognition 2001.

[8] Rainer Lienhart and Jochen Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection". IEEE ICIP 2002, Vol. 1, Sep. 2002.

[9] OpenCv, "https://opencv.org/links/".

[10] Shoaib Ehsan[1]*, Adrian F. Clark[1], Naveed ur Rehman[2] and Klaus D. McDonald-Maier[1],"Integral Images: Efficient Algorithms for Their Computation and Storage in Resource-Constrained Embedded Vision Systems," Sensors

[11] Kinh Tieu and Paul Viola, "Boosting Image Retrieval", IEEE Conference on Computer Vision and Pattern Recognition 2000.

[12] Constantine P. Papageorgiou, Michael Oren, Tomaso Poggio, "A General Framework for Object Detection".