A Study on Face Detection Using Viola-Jones Algorithm in Various Backgrounds, Angles and Distances

Tanmoy PAUL¹, Ummul Afia SHAMMI¹, Mosabber Uddin AHMED¹, Rashedur RAHMAN²
Syoji KOBASHI² and Md Atiqur Rahman AHAD¹,³

¹Department of Electrical and Electronic Engineering, University of Dhaka
²Graduate School of Engineering, University of Hyogo
³Department of Media Intelligent, Osaka University

Abstract: In recent years, research in image processing is very popular and demanding because of its diverse and humongous application. Face detection is one of the dominant mechanisms for security purpose, video surveillance, humane computer interaction, etc. Numerous algorithms have been proposed by many types of research regarding face detection and Viola-Jones algorithm is one of them. This paper deals with Viola-Jones algorithm for face detection in various positions and surroundings which include distance from the camera, background color, the angle of the object, etc. Further, this paper also performs an analysis in order to find the best settings that are appropriate for Viola-Jones algorithm.

Keywords: Face detection, Viola-Jones algorithm, Various background

1. Introduction

Human life is augmented by artificial intelligence, in which a machine exhibits intellectual functions. Computer vision tends to mimic humane vision [1]. For more than 20 years, research in this sector has been operated and therefore we can say that the current status of this field is progressive. Face detection is a very famous and attractive research topic since it has multifarious applications. One of the biometric authentication techniques is face recognition and face detection is the primary step of it [2]. Human beings are a gift with an amazing inborn potential to detect and recognize a person by just using his memory. Unfortunately, the computer does not have this function and that is where artificial intelligence appears.

Computer vision-based action recognition, face recognition, facial action unit recognition, eye gaze detection, tiredness understanding, emotion understanding, tracking, gait analysis, healthcare, etc. are very important active research areas [4-6,12,14,15,21,36]. Though recording ‘face’ relates to privacy issues, it is a very widely used scheme in numerous vision-based applications over a decade. By doing sophisticated programming and training, a machine can learn and duplicate humane activities. Visual emotion perception can be analyzed by face-scanning [3]. Face detection and then approaching various actions from detected faces are gaining in many applications [4,5,29,36].

The function of this face detector is to detect a face and track it wherever it goes where the input is taken from a video camera. Different approaches can be adhered [6]. An exact numerical description of faces and non-faces should be given to a computer in order to train the computer. These characteristics can be extracted with AdaBoost. It is an exceptional committee learning algorithm and uses the voting mechanism to form a strong classifier from the weak one [7]. Viola-Jones object detection algorithm was proposed by Viola and Jones in 2001 where techniques such as integral image and cascade make the algorithm highly efficient [8]. This algorithm first searches for vertical bright bands in an image which might be noses. Then it finds out horizontal dark bands which might be eyes. Finally, it looks for other general patterns associated with faces. These features individually may not suggest face strongly, but once these are detected one after another in a cascade, they are
strongly suggestive of the face. As long as the average time to detect face is concerned, this algorithm shows better performance than other face recognition techniques, such as Gabor features extraction and project profile analysis [9].

In this research, we will use this Viola-Jones algorithm in various surroundings and environmental conditions. If we consider a video or an image of a person where the background is excluded, that would be unrealistic. Therefore, a more pragmatic approach would be to take a video in a complex background where other things are taken into account. The recognition rate will differ in every case.

The paper is organized as follows. Various methods of face detection are discussed in Section 2. Section 3 discusses Viola-Jones algorithm. The experimental setup, the detection result, and its analysis are shown in Section 4. Section 5 concludes the paper with a summary.

2. Literature Review

In this section, we will have a look at existing face detection techniques. 3D facial detection is getting better and can be classified into four categories. They are presented in the following sub-sections.

![Figure 1. Various types of face-detection](image)

**2.1. Knowledge-based method**

This method exploits the human knowledge of the typical geometry of human face and facial organization [10]. Face localization can be achieved by using these methods. Yang and Huang [11] used a knowledge-based method in a hierarchical form to detect faces. There are three levels of rules in their rules. At the highest level, the window method is used for finding all possible face candidates. A window is scanned over the input image and a set of rules is applied at each location. A rule-based localization was proposed by Kotropoulos and Pitas [31] where the horizontal and vertical profiles of an input image are obtained.

**2.2. Feature invariant method**

Contrary to the knowledge-based method, feature invariant method finds structural features for face detection. Different structural features like facial local features, shapes, texture and skin color are being used [10]. A statistical model is then developed to depict their associations to certify the presence of a face. Sirohey [13] proposed a method in which the head is separated from the background clutter by employing an elliptical model of head. Through some preprocessing, it utilizes this information present in the edge map and an ellipse is then fitted to mark the borderline between the head area and the background. Augusteijn and Skujca [32] developed a method that used face-like texture to detect the presence of a face. They examine skin, hair and other features of the face. Textures were computed by exploiting the second-order statistical features. Sub-images of $16 \times 16$ were taken for this purpose.

**2.3. Template matching approaches**

Template matching approach makes use of correlation values. Here, at first, a function is used for a standard face pattern. After an input image is given, the correlation values along with the standard patterns are a measure for different parts like eyes, nose, face contour and mouth [33]. A multi-resolution template matching with color segmentation and region clustering was developed by Ping et al. [34]. Jin et al. [17] also use a template matching approach but in a different manner. Firstly, skin color information is used based on a luminance-conditional distribution model; then, morphological operations are used and followed by template matching. Here, for detecting a face in each skin-region rectangle, template matching approach is based on linear transformation.

**2.4. Appearance-based methods**

The basic difference between the template matching method and appearance-based method is that the later approach can learn from a large number of examples [4]. Detecting faces are similar to a two-class pattern classification problem, which are face and non-face. All face images are in the face class and the non-face class contains images that may portray anything but a face [6].
principal component analysis (PCA) along with Eigenface technique is an example of an appearance-based method and one of the most used algorithms. Mutelo et al. [18] developed two-dimensional reductions PCA (2D PCA) which represents data image matrices within fewer rows and columns. Jee et al. [19] utilize support vector machine (SVM) for detecting face region which is a linear classifier. Anand and Lawrance [20] proposed a face detection and recognition algorithm based on the left-right hidden Markov models (HMM). It also uses singular value decomposition (SVD) coefficients. The human face is segmented into seven facial regions and for choosing facial features, a reduced set of quantized SVD coefficients was trained. Jameel [35] used HMM and PCA coefficients as features and then applied discrete cosine transform (DCT). It then uses seven-states HMM to model face configuration. Hashemi and Gharahbagh [22] proposed a method for feature extraction that uses wavelet transform with eigen value of 2D wavelet coefficients matrix. Here, radial basis function neural network is used as a classifier. Gao and Lee [23] developed a method combining the affine scale invariant feature transform (SIFT) and probabilistic similarity, for face detection and recognition under a series of different viewpoints. Merging of both global and local features with the use of PCA and local binary pattern (LBP) has been developed by Sompura and Gupta [24]. It results in one vector which is used by multilayer perceptron (MLP). Alshebani et al. [25] presented an innovative method for feature extraction. The features are extracted after segmenting the facial regions of the nose, eyes, and mouth manually. For calculation of the local representations of these regions, Gabor transform of the maximum of these areas is extracted. And for the classification, the K nearest neighbor (KNN) method is used. Wu [26] uses horizontal component prior principle (HCPP) and to improve LBP operator, it exploits the regional directional weighted LBP (RDW-LBP).

3. Viola-Jones Method

Viola-Jones algorithm is the most famous algorithm for face detection. It is a framework that is used in real time and the training rate is very high. There are three major ideas in Viola-Jones face detection method: the integral image, AdaBoost classifier learning, and finally the attentional cascade structure.

3.1. Rectangular feature

Instead of pixels, Viola-Jones algorithm uses rectangle features. A finite quantity of training data is insufficient to encode ad-hoc domain knowledge. Rectangular features can solve this problem [27].

In Figure 2, two-rectangle feature means the difference between the summations of the pixels inside two rectangular zones. Consequently, three-rectangle feature means the summation of two external rectangular regions subtracted from the summation in a center rectangle. Difference between diagonal pairs of rectangles is indicated by four-rectangle feature [8].

![Figure 2. Rectangle features relative to enclosing detection window](image)

3.2. Integral image

At locations $x, y$, the integral image holds the sum of the pixels above and to the left of $x, y$ inclusive:

$$\text{i}(x, y) = \sum_{x', y' < y} i(x', y')$$  \hspace{1cm} (1)

The computation of integral images can be carried out in one pass over the original image by utilizing the following pair of recurrences:

$$s(x, y) = s(x-1, y) + i(x, y)$$  \hspace{1cm} (2)

$$\text{i}(x, y) = \text{i}(x-1, y) + s(x, y)$$  \hspace{1cm} (3)

Here, $s(x, y)$ means the cumulative row sum.

In rectangle D, with four array references, the sum of the pixels can be computed. These are: the summation of the pixels in rectangle indicated as A is assigned as the value of the integral image at location 1. A+B is the value at location 2. The value at location 3 is A+C, and A+B+C+D is considered as the value at location 4.

Computation of sum within D is done as $4 + 1 - (2 + 3)$. For any scale, six array references are used to compute a two-rectangle feature [8].
3.3. Learning classification functions

Variant AdaBoost is used for feature selection and classifier training purpose [28]. It combines a group of weak classification functions and forms a classifier which is stronger. To boost the weak learner, it is to solve a series of learning problems. Once the first round of learning is completed, the examples are re-weighted. The previous weak classifier classifies with less accuracy and the re-weighting process takes care of those. The final classifier is a strong one and takes the structure of a perceptron [30].

3.4. AdaBoost for feature selection

Given example images \( x_i \) and labels \( y_i = \{0, 1\} \), let us initialize weights according to the following equation:

\[
W_{t,1} = \frac{1}{2m} \frac{1}{2l} \quad (4)
\]

where \( m \) and \( l \) respectively mean the number of negative and positives. For \( t = 1, \ldots, T \), we normalize the weights:

\[
W_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{m} w_{t,j}} \quad (5)
\]

Afterward, we select the min-error classifier \( h_t \),

\[
\epsilon_t = \min_{f, p, \theta} \sum_i w_i \left| h(x_i, f, p, \theta) - y_i \right| \quad (6)
\]

If \( x_i \) is classified incorrectly, its weight is not modified. Otherwise, its weight needs to be adjusted according to the following equation:

\[
W_{t+1,i} = W_{t,i} \frac{\epsilon_t}{1-\epsilon_t} \quad (7)
\]

3.5. Final (Strong) classifier

Weak classifiers are linearly combined and this combination forms the final classifier. Performance of each classifier is used to weight it:

\[
C(x) = \text{sign} \left[ \sum_{t=1}^{T} \left( \log \frac{1-\epsilon_t}{\epsilon_t} \right) \left( h_t(x) \frac{1}{2} \right) \right] \quad (8)
\]

Note that for \( \epsilon_t = 0.5 \), the classifier \( t \) fails to participate in the combination.

It is easy to interpret the first two features for the purpose of face detection. The region of the eyes is the first feature selected by AdaBoost. This region is usually darker than the region of cheeks and nose. The second feature is the eyes which are often darker compared to the bridge of nose [8].

3.6. Attentional cascade

The cascaded classifier is more efficient because it is smaller. Such a classifier rejects most of the negative subwindows and detects the majority of the positive ones. In order to discard a major portion of subwindows, simpler classifiers are used. More complex classifiers are used to gain low false positive rates.

![Figure 5. Schematic representation of the detection cascade (here, T: True; F: False)](image)

For a cascade of a classifier that is trained, the false positive rate is [27]:

\[
F = \prod_{i=1}^{K} f_i \quad (9)
\]

where \( F \) denotes the false positive rate of the entire cas-
cascade, the number of classifiers is denoted by $K$, and $f_i$ is the false positive rate of the $i$th classifier. The rate of detection is:

$$D = \prod_{i=1}^{K} d_i$$

(10)

where $d_i$ is the detection rate of the $i$th classifier on the examples.

The cascade training algorithm is as follows. While global false positive (FP) rate is not met, $n$ takes the value 0. The value of $n$ is increased by 1 so that it is $n + 1$. Then the classifier is trained from $n$ features with AdaBoost and it is evaluated on the validation set. After that, the classifier’s threshold is decreased until the detection rate becomes at least $d$. Then again the value of $n$ is increased until a classifier of FP rate less than $f$ is found. Finally, the classifier is added to cascade and future classifiers are trained on false positives.

### 3.7. Final training

In the final training of Viola-Jones method, Adaboost training was used and 4916 faces along with 10,000 non-face sub-windows were used to train the two, five and first twenty-feature classifiers. Non-face sub-windows were gathered from the images, which did not contain faces. This was selected by sub-windows as random manner, from an image set comprising 9500 images. To train different classifiers, a number of diverse sets of non-face sub-windows were utilized to make sure that they were independent. Another reason to do so is to ensure that they did not use the same features [8].

At multiple locations and scales across the image, the final detector is scanned. Instead of scaling the image, the detector itself is scaled. Window shifting, by some number of pixels, allows the detector to be scanned. During training, in order to minimize the different effects of lighting conditions, sub-windows were variance normalized. The same measure is taken during detection. To get a single detection by combining all the overlapping detections, the detected sub-windows are post-processed [4].

### 4. Experimental Result and Analysis

In this paper, we have created our own data set. We have chosen 3 types of the background while making it. They are a white background, black background, noisy background. We have also varied the distance between the camera and the face. The distances between the face and the camera are as follows: 5 feet, 8 feet, 12 feet and 15 feet away from the camera. We have varied the angle of the face in the data set as well. The angels are 30-degree, 60-degree, 90-degree angle apart from the camera. We have used Xiaomi Redmi Note 4 phone which is a 13 Megapixel camera. For programming, we have used MATLAB 2015. The result of the detected face is as follows.

### 4.1. In black background

![Figure 7. (a) 5 feet (b) 8 feet (c) 12 feet (d) 15 feet away from the camera](image)
4.2. In white background

Figure 8. (a) 5 feet (b) 8 feet (c) 12 feet (d) 15 feet away from the camera

4.3. In noisy background

Figure 9. (a) 5 feet (b) 8 feet (c) 12 feet (d) 15 feet away from the camera

It is seen that in Figure 7, the accuracy of detecting the face using Viola-Jones algorithm is 100% because there is a high contrast between the dress color of the object and the background. In Figure 8 the dress color of the object and the background is almost the same. In this case, the method shows poor accuracy when the object is 12 feet and 15 feet away from the camera. This is also to be mentioned that the camera quality is poor and the object’s face is not clear in Figures 8(c) and 8(d). In Figure 9, the background is noisy and there is another object in the background. The face detection accuracy is good in Figures 9(a) and 9(b) but in the later figures, the face is not detected accurately. Another reason other than the dress color of the object is that the floor in the figures lights deep color variation.

4.4. Variation of angles

Figure 10. Variation of angles from the camera (30°, 60°, 90°)

4.5. In black background (8 feet away)

Figure 11. The angle between the face and the direction of the camera is (a) 30 degree (b) 60 degree (c) 90 degree.

4.6. In white background (8 feet away)

Figure 12. The angle between the face and the direction of the camera is (a) 30 degree (b) 60 degree (c) 90 degree.
4.7. In white background (8 feet away)

In these figures, the object is 8 feet away but the angle between the face and the direction of the camera is 30 degree, 60 degree, and 90 degrees. In Figure 11, again the contrast with the background is high since the background color is black, that is why the detection is perfect. In the following Figure 12, the detection is also accurate in these figures because the object is 8 feet away from the camera. In the next figure, Figure 13, where the angle between the face and the direction of the camera is 90 degree, detection is not accurate.

Figure 13. The angle between the face and the direction of the camera is (a) 30 degree (b) 60 degree (c) 90 degree.

Table 1. The experimental result of face detection at various distances and angles

<table>
<thead>
<tr>
<th>Variation is distance</th>
<th>Black Background</th>
<th>White Background</th>
<th>Noisy Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 feet</td>
<td>Correct detection</td>
<td>Correct detection</td>
<td>Correct detection</td>
</tr>
<tr>
<td>8 feet</td>
<td>Correct detection</td>
<td>Correct detection</td>
<td>Correct detection</td>
</tr>
<tr>
<td>12 feet</td>
<td>Correct detection</td>
<td>False detection</td>
<td>False detection</td>
</tr>
<tr>
<td>15 feet</td>
<td>Correct detection</td>
<td>False detection</td>
<td>False detection</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variation in angle (8 feet away)</th>
<th>Black Background</th>
<th>White Background</th>
<th>Noisy Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>30°</td>
<td>Correct detection</td>
<td>Correct detection</td>
<td>Correct detection</td>
</tr>
<tr>
<td>60°</td>
<td>Correct detection</td>
<td>Correct detection</td>
<td>Correct detection</td>
</tr>
<tr>
<td>90°</td>
<td>Correct detection</td>
<td>Correct detection</td>
<td>False detection</td>
</tr>
</tbody>
</table>

5. Conclusion

Face detection is a difficult task under various constraints and distances. In this paper, we have explored face detection in various situations. After this research, we concluded that Viola-Jones method works well when the contrast is high between the background and the subject is high. The face detection is not accurate for noisy background. The detector can detect faces when the face of the subject creates an angle of 60 degrees or less with the camera. This performance of this method starts to decrease for the higher angle. Overall, Viola-Jones method is by far most widely used and an efficient method of detecting a face. The same method can be used to detect other objects. In that case, new training is needed.
References


[23] Alshebani, Q., Premaratne, P. and Vial, P.: A hybrid feature extraction technique for face recognition,


Tanmoy PAUL
He has received B.S. in Electrical and Electronic Engineering from University of Dhaka in 2016. Currently, he is doing his M.S. from the aforementioned department.

Ummul Afia SHAMMI
She has received B.S. in Electrical and Electronic Engineering from University of Dhaka in 2016. Currently, she is doing her M.S. from the aforementioned department.

Mosabber Uddin AHMED
Mosabber Uddin Ahmed received the B.Sc. (Hons.) and M.Sc. degree in Applied Physics and Electronics from University of Dhaka, Bangladesh, in 2000 and 2003 respectively. He received the M.Sc. degree in Communications and Signal Processing from Imperial College London in 2006 and Ph.D. degree in signal processing from the same institution in 2012. Currently, He is an associate professor at the Department of Electrical and Electronic Engineering in University of Dhaka, Bangladesh. His research interests include biomedical signal processing, embedded system design, and complexity science. He was also the recipient of the Commonwealth Academic Fellowship in 2015.

Rashedur RAHMAN
Rashedur Rahman received his B.S. (2018) in Electrical and Electronics Engineering from International University of Business Agriculture and Technology, Bangladesh. Currently, he is working as a researcher in Advanced Medical Engineering Center (AMEC) of University of Hyogo, Japan. His research interests include image processing, signal processing and machine learning.
Syoji KOBASHI

Syoji Kobashi received Dr. of Engineering from Himeji Institute of Technology (2000). He is currently a professor and the director of advanced medical engineering center since 2017, and an assistant dean of Graduate School of Engineering since 2018 at University of Hyogo. His research interests include artificial intelligence in medical engineering. He is the editor at large of autosoft journal, and a senior member of IEEE.

Md Atiqur Rahman AHAD

Md Atiqur Rahman Ahad (Senior Member, IEEE) is a Professor, University of Dhaka (DU); and currently as a Specially Appointed Associate Professor, Osaka University. He works on computer vision, imaging, IoT & healthcare. He did B.Sc. (Honors) & Masters (DU), Masters (University of New South Wales), Ph.D. (Kyushu Institute of Technology), JSPS Postdoctoral Fellow and Visiting Researcher. He published two books (in Springer). Others: Editorial Board Member, Scientific Reports, Nature; Associate Editor, Frontiers in ICT; Editorial Board Member, Encyclopedia of Computer Graphics and Games, Springer; Assoc.-Technical Editor (former), IEEE ComSoc Magazine; Editor-in-Chief, IJCISP http://cennser.org/IJCISP, IIEI, IJE; General Chair, 7th ICIEV, http://cennser.org/ICIEV; 2nd IVPR, Publication Chair, IEEE SMC Conf. 2018; Publication Chair, 17th IFSA Annual Conf. 2017; Guest-Editor: Pattern Recognition Letters, Elsevier; JMUI, Springer; JHE, Hindawi; IJICIC; Member: OSA, ACM, etc. He volunteers some societies in BD/JP. More: http://aa.binbd.com.