How to Cite:
Abdulhussien, D. M., & Saud, L. J. (2022). An evaluation study of face detection by Viola-Jones algorithm. *International Journal of Health Sciences*, *6*(S8), 4174–4182. https://doi.org/10.53730/ijhs.v6nS8.13127

An evaluation study of face detection by Viola-Jones algorithm

Dina M. Abdulhussien
Department of Control and Systems Engineering, University of Technology- Iraq
Corresponding author email: mvzyo54980@gmail.com

Laith J. Saud
Department of Control and Systems Engineering, University of Technology- Iraq

***Abstract*---**Face detection technology is an essential step in almost all face related analysis applications such as face feature extraction, face alignment, face verification, face identification, face parsing, face recognition, age recognition, and gender classification. Numerous algorithms were introduced for face detection, one of which is the Viola-Jones algorithm being introduced in 2001. This algorithm is still widely used due to its simplicity and ability of detection in real-time with relatively high accuracy and low computational power requirements compared to other recent algorithms such as deep learning based algorithms. In this paper, Viola-Jones algorithm is implemented and evaluated through different tests. And its strengths, limitations, and affecting factors are provided according to the obtained results. This paper concentrates on the algorithm limitations and the reasons of these limitations, and suggests some solutions if possible. This can help in enhancing the algorithm performance by increasing the detection accuracy or reducing the time taken for detection or training…etc.

***Keywords*---**face detection, viola-jones, Haar-like features, integral image.

Introduction

Viola-Jones face detector is a feature-based approach [1], which was introduced in 2001 by Paul Viola and Michael Jones [2]. It was the first face detection algorithm used in real-time. It was known for its capability of processing images extremely rapidly while achieving high detection rates [3]. Face detector is essential to almost all recent important applications based on face analysis, such as face image enhancement [4] , face recognition [5][6], face identification [7], face verification, face alignment [8], face feature extraction [9], face emotion

recognition, and gender classification. One of the popular applications used today in mobile devices is the snapchat (snapchat application apply filters to face but first it has to find the face, and this is obtained by using Viola-Jones). Viola jones has many advantages, such as relatively high detection speed and accuracy, low computational power requirements, ability to use in real-time, it can be learned to detect any object not just human face, and it is an ensemble classifier that consists of several weak classifiers. Ensemble classifiers are considered to be more stable, predict better, and achieve better performance than a single classifier. But it still faces several challenges such as difficulty in detecting non-frontal faces, faces with unclear borders, detecting some animal and sketched faces falsely, images with too high resolution or images with too low resolution (the detection window is unable to recognize face features clearly), faces with abnormal expressions, partially occluded faces, image illumination variations, and blurred image. In this paper some tests that focus on some failure cases in face detection by viola jones are considered with discussing the reasons and suggesting solutions if possible.

Related works

- In 2014, Wang [10] analyzed viola jones face detector and indicated that three important parts are behind the accurate and fast detection. These are: the integral image used for rapid feature calculation, Adaboost learning algorithm used for selecting small set of best features from larger set and a cascade classifier used for efficient computational resource allocation.
- In 2015, Egorov et. al. [11] studied the effect of five parameters on the viola-jones operation. These parameters are: 1) Cascade used for searching objects. 2) Scaling coefficient of the change in the scanning window size when going from one algorithm iteration to another. 3) Number of neighbors. 4) Initial size of the scanning window. 5) Maximum size of the scanning window. They concluded that, the algorithm operating time strongly varies depending on parameters when using mentioned cascades. And the highest accuracy obtained at the least operating time is provided by LBP cascades.
- In 2016, Dabhi et. al. [12] implemented the viola jones method for face detection in real-time. They used the static data base of CMU PIE which contains (106) images with various backgrounds and lightning conditions and the result efficiency obtained was 87%.
- In 2017, Vikram [13] focused on Viola-Jones algorithm capability in human facial parts detection. The face, eyes, nose and mouth is detected in a random set of samples and also tested. They indicated the possibility of building on their algorithm to detect various actions of a human.

The work concept of Viola-Jones algorithm

Viola-Jones algorithm is composed of four parts; Haar-like features, integral image, Adaboost learning algorithm, and the cascade classifier.

1) Haar-like features: Viola-Jones extract Haar-like features to detect relevant features in human face such as (edge or line) [2]. They owe their name to their intuitive similarity with Haar wavelets (haar wavelets is Haar basis

functions that encode differences in average intensities between different regions). Haar-like features are determined by finding the difference between the sums of pixel values in adjacent rectangular regions [10]. The set of rectangle features provides a rich image representation, which supports effective learning. Viola-Jones uses three types of Haar-like features:

- Two-rectangle feature: This is determined by finding the difference between the sums of the pixel intensities in two rectangular regions. These regions are similar in size and shape and they are adjacent either in vertical or horizontal direction. The two rectangle features are usually used to detect edges.
- Three-rectangle feature: This is determined by finding the sum within two outside rectangles then subtract it from the sum in the middle rectangle. This rectangle features are usually used to detect lines.
- Four-rectangle feature: This is determined by finding the difference between diagonal pairs of rectangle regions.

  Generally, the value of each Haar-like rectangle feature is compared to a learned threshold that separates non-targeted objects from targeted objects. The searching process is repeated in different locations and sizes, and the resulted differences are used to categorize the tested subsections of the image window. In this way, face landmarks such as (eyebrows, lips, nose, …etc.) are detected. The number of Haar-like rectangle features inside the image sub-window of size (24 x 24) pixels is more than (160,000). This enormous calculations put a limit on using the algorithm in real-time. However, computing Haar-like rectangle features using integral image representation, will reduce computation time and complexity to a very large extent.

2) Integral image: Also called sum-area table, is a very important contribution introduced by Viola and Jones which enables computing the large number of Haar-like features very rapidly and at many scales. The sum-area table can be determined from the input image by using only a few operations per pixel. Each pixel in the integral image contains the cumulative sum of the corresponding input image pixel with all pixels above and to the left of it. Once the integral image is created, each of the Haar-like features can be computed at any scale or location in constant time. This is because the number of array references required to determine any rectangle region of Haar-like features using the integral image is constant, which is only four references. i.e. the integral image reduces computation complexity from O (m ×n) to O (1), where m × n represents the resolution of the image, and this is a huge reduction [3].
3) Adaboost learning: Adaboost learning algorithm, short for Adaptive Boosting, was introduced in 1995 by Freund and Schapire. In Viola-Jones algorithm, Adaboost is used both to select the features and to train the classifier [2]. Adaboost selects a small number of important features (weak classifiers) from a larger set of features to produce a very efficient classifier (strong classifier). The total number of Haar-like features inside any detection window is very large. To ensure fast classification, the learning process must focus on a small number of best features and exclude a large majority of available features. In the Viola-Jones algorithm, the first feature selected by Adaboost determine the difference in intensities between the eyes region and the region across the upper cheeks. The feature take advantage of the

observation that the eyes region is often darker than the below region across the upper cheeks. While the second feature determines the difference between the intensity of the eye regions and the intensity across the bridge of the nose [2].

4) Cascade classifier: It is a method to combine increasingly more complex classifiers in a cascaded way, which quickly discards background regions of the image while spending more computation on promising face-like regions. Strong classifiers are formed into a binary classifier, where positive matches are passed to the next classifier and negative matches are rejected and the algorithm exits computation [3]. A series (cascade) of classifiers is applied to every sub-window. The initial classifiers use less number of features and eliminates a large number of negative examples with very little processing. Subsequent classifiers use more features and eliminate additional negatives but require more computation. The number of sub-windows is reduced radically after several processing stages. The use of cascade classifiers enhance the detection performance while radically reduces the time required for computation. Mostly, each stage in the cascade classifiers reduces the false positive rate and increase the detection rate [2].

Viola jones algorithm work consists of two stages: training stage and detection stage. The outcome of the training stage is a classifier used later in the detection stage to detect face from non-face. The aim of this paper is to study and evaluate the ability of Viola-Jones algorithm on face detection in different situations.

Training stage

Training starts with feeding data to the algorithm, and it requires a large amount of face images to be able to extract features in different forms. Firstly, each input image will be converted to an integral image representation. Then Haar-like features are calculated (in several sizes and locations) efficiently using the integral image. The number of the produced Haar-like features is very large (more than 160,000 within a sub-window of 24 x 24 pixels) [3]. After that, Adaboost is used to select a small number of the best features (which nevertheless have significant variety) from the larger number of available features. Lastly, cascade of classifiers are used to eliminate as many negatives as possible, while detecting almost all positive instances. The output of the algorithm is saved as an XML file and used later to detect faces.

Detection stage

Viola-Jones was first introduced to detect frontal faces, so it is able to detect frontal faces better than tilted or rotated faces. Before the face detection process is started, the image is converted to a gray color, because it is easier to work with and there's less data to process. After that, the trained cascade classifier is used to detect the faces. This is done using a detector sub-window, which will scan the whole image at multiple scales and locations to search for a face within it. The sub-window moves a step to the right after going through every pixel in the image. With smaller steps, a number of sub-windows detect Haar-like features and the data of all of those sub-windows are put together, which helps the algorithm determine where the face is [3].

Training dataset

Viola and Jones mentioned in their paper [3] that they used a training dataset that consists of two image sets (positive and negative). The positive set is a collection of face images (4916 hand labeled faces are scaled and aligned to a base resolution of 24 x 24 pixels). The face images were randomly selected from the World Wide Web. While the negative set comes from (9544) images, which were manually inspected and it was found that they do not contain any faces, not even partially. The training faces are aligned roughly by placing a bounding box around each face just above the eyebrows region and about halfway between the mouth and the chin region. This bounding box was then enlarged by 50% and cropped and scaled to (24 x 24) pixels. In this paper, Viola-Jones algorithm is implemented on detecting faces in five different tests:

The first test

Input images used in this test consist of (25) animal (non-human) faces. The resolution of these images is varying from (187 x 270) to (2500 x 1668) pixels. In this test, Viola-Jones is supposed to fail in detecting these faces (as they are not human faces). This test is repeated more than once to check results repeatability. The results showed that monkey faces are sometimes detected as shown in Figure (1).

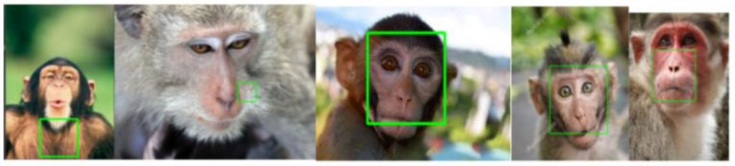

Figure (1): Falsely detected monkey faces

This test indicates that, some monkey faces are falsely detected or located because of the similarity between the human features and monkey face features in (location, size, and distance between each other's). And this is considered as a flaw in the algorithm. This could be solved by increasing number of features used in training the classifier which helps in distinguishing human faces from none.

The second test

The input images in this test consist of (4) images for elder human faces. Images resolutions start from (183 x 275) to (600 x 877) pixels. We found that the faces are not detected or falsely detected. After that we repeated the same test on the same images but after reducing their size. It is founded that, some elder people faces are detected correctly. The main reason could be due to the large face size compared to the detecting window size (around 24 x 24 pixels), so the detecting window is unable to detect the face features. Another reason could be the several wrinkle lines appear on their faces which confuse the process of extracting face features correctly.

Figure (2): faces not or falsely detected.

Figure (3): some faces detected after reducing size.

The third test

Input images in this test consist of (4) human faces without borders as shown in the Figure (4). Borders are deleted here by Adobe Photoshop. The algorithm failed in detecting these faces. It is concluded from this test that the existence of face borders such as chin, cheek, and hair lines is vital to detect a face.

Figure (4): Faces without borders

The fourth test

In this test, input images are (17) sketched faces, some of them shown in figure (5). And (10) faces are detected. Viola-Jones should detect only real not sketched faces. Detecting sketched faces can be avoided by applying skin-color classifier before applying cascade classifier. So only human faces with normal skin color is passed as input to the cascade classifier. On the other hand, detecting sketched faces is useful in telling us which face features are exactly used by viola jones within detection process.

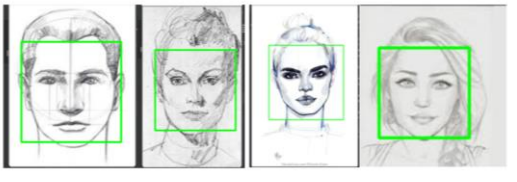

Figure (5): some of the detected sketched faces

The fifth test

Input faces are (14) faces with different and abnormal expressions. Two faces are not detected when implementing viola jones, as shown in Figure (6). This is could be because the used viola jones classifier does not trained on such face features.

Figure (6): Faces with abnormal expressions

Some of Viola-Jones algorithm strengths and limitations are listed below according to Viola-Jones papers and our tests.

Strengths of the Viola-Jones algorithm

Viola-jones algorithm was firstly introduced in 2001 for face detection, but it still quite powerful due to several reasons. Some of them are given below:

- Viola-Jones detector processes images with relatively high detection accuracy.
- The algorithm is efficient in detecting faces in real-time.
- Viola-Jones algorithm is first introduced for face detection, but it can be used for any object detection.
- Viola-Jones algorithm requires a minimal memory requirements and it is simple to implement. Hence, it is used in many applications for embedded devices and mobile devices.
- Viola-Jones requires about (20,000) training samples, and this is far less compared to algorithms based on deep learning, which require millions or even billions of training samples.
- Viola-Jones is a strong ensemble classifier which consists of many weak classifiers.

Limitations of the Viola-Jones algorithm

- Detection accuracy declines for faces of elder people. The reason may be due to the wrinkle lines which may confuse face feature extraction, or it could be due to the large face size compared to the detecting window size (around 24 x 24 pixels), so the detecting window is unable to detect the face features.
- Moreover, a low detection accuracy is achieved for faces with abnormal expressions. Because the face features may not be discriminated from each other.
- It works well in detecting frontal faces, and to some extent may detect tilted or rotated faces, but cannot detect profile (side-viewed) and upside-down faces.
- Preparing appropriate quality and quantity of images to be used for the training process can be exhaustive.

- Training time is very long, and it can take weeks.
- It is not clear how to select few features in the first cascades or how many samples are needed and how to gather a good training set for training a cascade.
- The detection by Viola-Jones is restricted to binary classification (existence, none existence) of an object.
- The face detection by Viola-Jones algorithm is highly prone to false-positive detections.

Some of the factors affecting the Viola-Jones algorithm performance

- Face borders (chin and cheek line, and hair line) have a noticeable effect on detection.
- Image illumination variation is important and it affects the detection process.
- The size and quality of dataset used for training the Viola-Jones algorithm is very important factor on its performance in the detection process. It requires a large number of face samples and a wide variety of face conditions such as several face (poses, directions, locations, scales…etc).
- Negative images used for training should not contain any faces, not even partially.
- Partial occluded faces cannot be detected.
- Image low resolution degrades detection capability.
- Large face size compared to the size of the detection window (around 24 x 24) pixels, can cause false detection or no detection at all.
- Blurred or damaged image affects so badly the detection process.

Conclusion

Although many face detection techniques are introduced after the Viola-Jones algorithm and some of them achieved better accuracy, but Viola-Jones is still widely used today in many applications such as mobile applications due to its relatively high detection accuracy in detecting frontal faces in real-time. Moreover, the viola-jones parameters is about (50k) compared to several millions of parameters for deep learning-based algorithms such as typical CNN, this enables viola jones to be used in devices with limited computational power such as embedded systems and mobile devices. This paper focus on some of viola jones difficulties appear in the detection process when using different image conditions such as animal (non-human) faces, faces that are large in size compared to the detection window, faces with unclear borders, faces with abnormal expressions, and sketched (not-real) faces. We suggest to improve the detection accuracy of viola jones classifier by training the cascade classifier on more features and this obtained by increasing number of cascade stages.

References

1. A. A. Kerim, R. F. Ghani, and S. A. Mahmood, “Robust alignment of salient facial regions for recognition of 3-D partial faces scans,” *2014 2nd Int. Conf. Electron. Des. ICED 2014*, no. January, pp. 73–78, 2011, doi:

10.1109/ICED.2014.7015774.
2. A. D. Egorov, A. N. Shtanko, and P. E. Minin, “Selection of Viola–Jones algorithm parameters for specific conditions,” *Bull. Lebedev Phys. Inst.*, vol. 42, no. 8, pp. 244–248, 2015, doi: 10.3103/S1068335615080060.
3. F. F. A. Bashra Kadhim Oleiwi, “Smart E-Attendance System Utilizing Eigenfaces Algorithm,” *Iraqi J. Comput. Commun. Control Syst. Eng.*, vol. 18, no. 1, pp. 56–63, 2018, doi: 10.33103/uot.ijccce.18.1.6.
4. G. Filters, “Face Image Enhancement using Wavelet Denoising and Gabor Filters,” *IRAQI J. Comput. Commun. Control Syst. Eng.*, vol. 16, no. 1, pp. 104–117, 2016.
5. G. Nauman, M. K. Dabhi, and B. K. Pancholi, “Face Detection System Based on Viola - Jones Algorithm,” *Int. J. Sci. Res.*, vol. 5, no. 4, pp. 62–64, 2016, doi: 10.21275/v5i4.nov162465.
6. K. Hasan, S. Ahsan, Abdullah-Al-Mamun, S. H. S. Newaz, and G. M. Lee, “Human face detection techniques: A comprehensive review and future research directions,” *Electron.*, vol. 10, no. 19, 2021, doi: 10.3390/electronics10192354.
7. P. Viola and M. Jones, “Robust Real-time Object Detection Paul Viola February,” pp. 1–25, 2001.
8. P. Viola, M. Jones, and M. Energy, “Robust Real-Time Face Detection Intro to Face Detection2004,” *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
9. S. Hamandi, A. M. Rahma, and R. Hassan, “Multi-Spectral Hybrid Invariant Moments Fusion Technique for Face Identification.” The International Arab Journal of Information Technology, Vol. 18, No. 3A, Special Issue 2021, Computer Science Department, University of Technology, Iraq, p. 9, 2021.
10. S. I. Mohammed, N. A. Jaafar, and K. M. Hussien, “Face Recognition Based on Viola-Jones Face Detection Method and Principle Component Analysis (PCA),” *Iraqi J. Comput. Commun. Control Syst. Eng.*, vol. 18, no. 3, pp. 52–59, 2018, doi: 10.33103/uot.ijccce.18.3.6.
11. S. L. Galib, F. S. Tahir, and A. A. Abdulrahman, “Detection Face Parts in Image Using Neural Network Based on MATLAB,” *Engineering and Technology Journal*, vol. 39, no. 1B. pp. 159–164, 2021, doi: 10.30684/etj.v39i1b.1944.
12. V. K and Dr.S.Padmavathi, “Facial Parts Detection Using Viola,” *Int. Conf. Adv. Comput. Commun. Syst. (ICACCS -2015)*, pp. 1–4, 2017.
13. Y.-Q. Wang, “An Analysis of the Viola-Jones Face Detection Algorithm,” *Image Process. Line*, vol. 4, pp. 128–148, 2014, doi: 10.5201/ipol.2014.104.