IOP Conf. Series: Materials Science and Engineering **732** (2020) 012038 doi:10.1088/1757-899X/732/1/012038

Comparison of Viola-Jones Haar Cascade Classifier and Histogram of Oriented Gradients (HOG) for face detection

C Rahmad[1,*], R A Asmara[2], D R H Putra[2], I Dharma[1], H Darmono[3] and I Muhiqqin[2]

[1] Information Technology Department, State Polytechnic of Malang. Jl. Soekarno-Hatta No. 9, Malang 65141, Indonesia.
[2] Electrical Engineering Department, State Polytechnic of Malang. Jl. Soekarno-Hatta No. 9, Malang 65141, Indonesia.
[3] Digital Telecommunications Network Department, State Polytechnic of Malang. Jl. Soekarno-Hatta No. 9, Malang 65141, Indonesia.

*cahya.rahmad@polinema.ac.id

Abstract. Human face recognition is one of the most challenging topics in the areas of image processing, computer vision, and pattern recognition. Before recognizing the human face, it is necessary to detect a face then extract the face features. Many methods have been created and developed in order to perform face detection and two of the most popular methods are Viola-Jones Haar Cascade Classifier (V-J) and Histogram of Oriented Gradients (HOG). This paper proposed a comparison between VJ and HOG for detecting the face. V-J method calculate Integral Image through Haar-like feature with AdaBoost process to make a robust cascade classifier, HOG compute the classifier for each image in and scale of the image, applied the sliding windows, extracted HOG descriptor at each window and applied the classifier, if the classifier detected an object with enough probability that resembles a face, the classifier recording the bounding box of the window and applied non-maximum suppression to make the accuracy increased. The experimental results show that the system successfully detected face based on the determined algorithm. That is mean the application using computer vision can detect face and compare the results.

1. Introduction

Simultaneous and automatic face detection in images, in real-time video, or offline video are three of the most studied topics in computer vision before face tracking, landmark detection, and face recognition. As the information on human identity, human faces are unique and cannot be replicated [1]. Face detection has been used for various purposes, such as supervision, identification of people, intelligent environments and robotics [2]. Thus, real-time and high detection accuracy are crucial factors. Face detection is a basic step for personal identification, monitoring system, criminal law, and human and computer interaction. The ever-evolving era demands the development of more accurate technology. More specifically, many problems in the field of technology and criminal security require identification of facial classification in solving problems [2]. In this paper, Viola-Jones Haar cascade classifier (V-J) and Histogram of Oriented Gradients (HOG) are used for face detection because both of these methods have a good level of accuracy in selecting an object. The detection of the face process

digital images to find out the characteristics of face. The Identified faces have different perspectives such as colors, sizes, and shapes and facial features in various sizes.

Furthermore, developing a face detection system has been developed to determine the object that resembles the face. In this work, the dataset obtained from our dataset, labeled Faces in the Wild (LFW), and Wider Face dataset with a high degree of variation of conditions like scale, pose, occlusion, expression, makeup, background, and illumination. Types of conditions used are shown in Figure 1 and the characteristics of datasets which include the number of faces and conditions are shown in table 1.

Figure 1. Type of condition.

Table 1. Characteristics of training datasets.

Variation of Condition	Number of Data Training	Number of Data test
Scale	20	8 Single Face and 7 Group
Occlusion	20	8 Single Face and 7 Group
Make Up	20	8 Single Face and 7 Group
Pose	20	8 Single Face and 7 Group
Expression	20	8 Single Face and 7 Group
Illumination	20	8 Single Face and 7 Group

This system uses the test dataset which will be processed using V-J and HOG, by using python the system can detect the face to compare the two methods.

2. Literature review

2.1. Computer vision

Computer Vision is a field of study that seeks to develop techniques to help the computer understand the content of digital images such as videos and photographs. Also, computer vision is automatic processing that consists of several large processes that are used for image processing, image acquisition, recognition, and decision-makers. An object image can be interpreted so that humans can find objects that appear in the eye so that a decision can be taken from the results of the interpretation [3].

2.2. *Viola-Jones*

V-J face detection algorithm scans an image with a window looking for features of a human face. If these features are found and have a particular value as a face, then the particular window of the image is estimated to be a face. To solve a case with different sized faces, the window scaled with the repeated process for each image. Reducing the number of features each window must be checked, the window passed through several different stages, which early stages had fewer features to checked and easier to pass while later stages have more features and more selective [4]. The features calculated for each stage then accumulated, If the accumulated feature didn't pass the threshold, it means the stage is failed and

the current window is estimated to not contain a face. To make it easier to understand the algorithm, some term needs to be defined including features, integral image, and stage cascade.

2.3. Features

Features named Haar classifiers are used in the V-J algorithm to detect features of a face. Haar features are used in computer vision to classified the intensity of pixels in a region with a traceable manner. Haar features represented as rectangle regions of the image, and the classifiers are composed of two or three rectangle features to continuously scanned for features in the window. An example of Haar features is showed in Figure 2. For more details of these features, we refer and the original work by Viola and Jones.

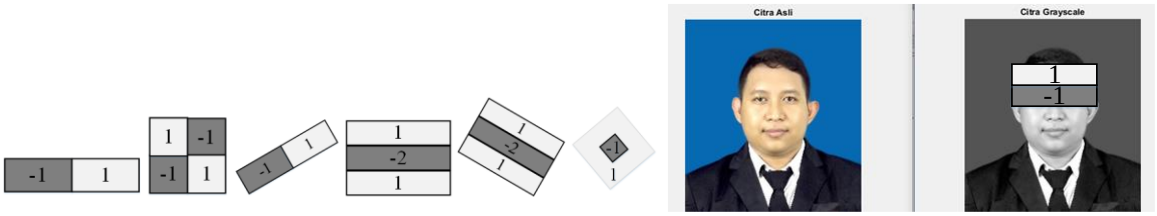

Figure 2. Haar feature.

2.4. Integral image

Compute the integral Image as a pre-processing step to avoid the calculation process in the sum excessively. Figure 3 shows how integral Image is calculated from pixel values. In an integral image, the value of each point is the sum of all pixels above and to the left, including the target pixel, then calculate the sum of pixels in the orange rectangle. Using formula D - B - C + A, the value of integral image is 113 – 81 – 42 + 20 = 41.

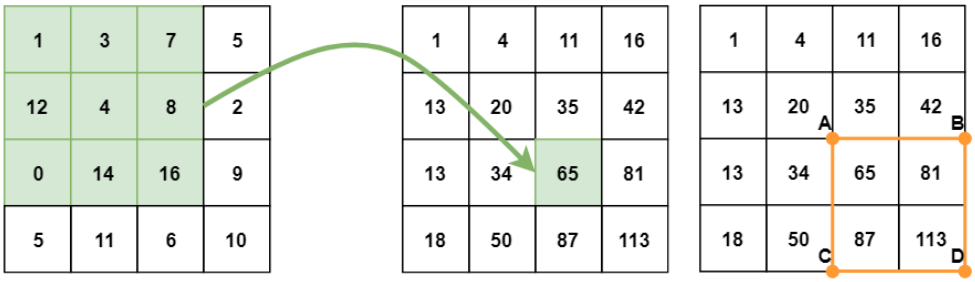

Figure 3. Integral image.

2.5. Adaptive Boosting (AdaBoost)

AdaBoost (AD) is a technique used to combine a collection of weak classifiers. The process of AD produced a strong classifier. One Integral Image is a weak learner, by combining many integral images by using AD, it creates robust classifications to determine facial features in the window area. The process of AD shown in Figure 4, the process goes from left to right. The missed blue samples are given more importance, which indicated by size. The second classifier captures the bigger blue circles and the misclassified orange circles are given more importance, while the others are reduced. The next step the third classifier captures the remaining orange circles, and the final strong classifier combines all three weak classifiers [5].

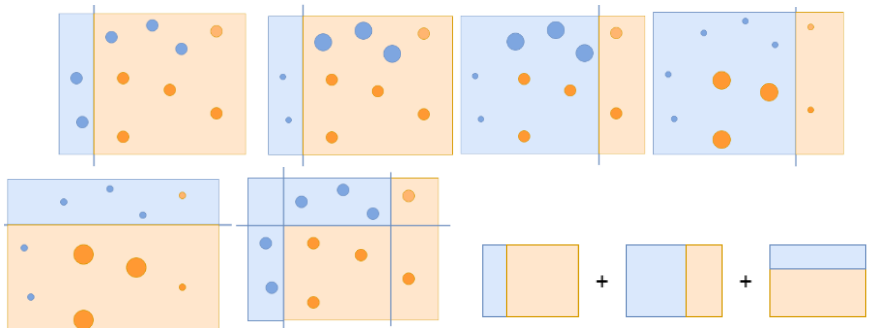

Figure 4. Adaptive Boosting.

2.6. Cascade classifier

Cascade classifier is the process of organizing a set of features in a multilevel classification form. There are at least three classifications to determine whether or not there are facial features in the selected feature area. In the first stage classification filter, each sub-image will be classified using one feature, if the value of the feature on the filter does not meet the expected criteria, it will be rejected. The algorithm then moves to the next sub window and calculates the value of the feature, if the results are following the desired threshold then it will proceed to the second filter stage until the number of sub-windows that pass the classification decrease until it approaches the detected face image. Figure 5 is a filtering process that is passed by each classifier.

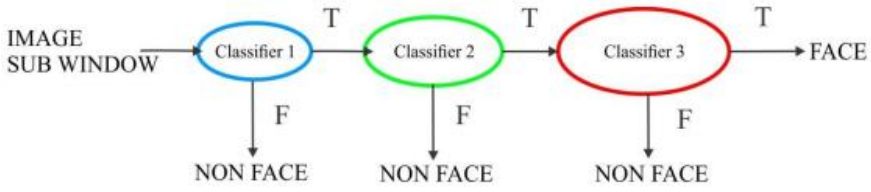

Figure 5. Cascade classifier.

2.7. Histogram of Oriented Gradients (HOG)

HOGs are a feature descriptor that has been successfully used for object and pedestrian detection, represented an object as a single value vector as opposed to a set of feature vectors where each represents a region of the image, computed by sliding window detector over an image. HOG descriptor is computed for each position, while the scale of image adjusted to get a HOGs feature [6].

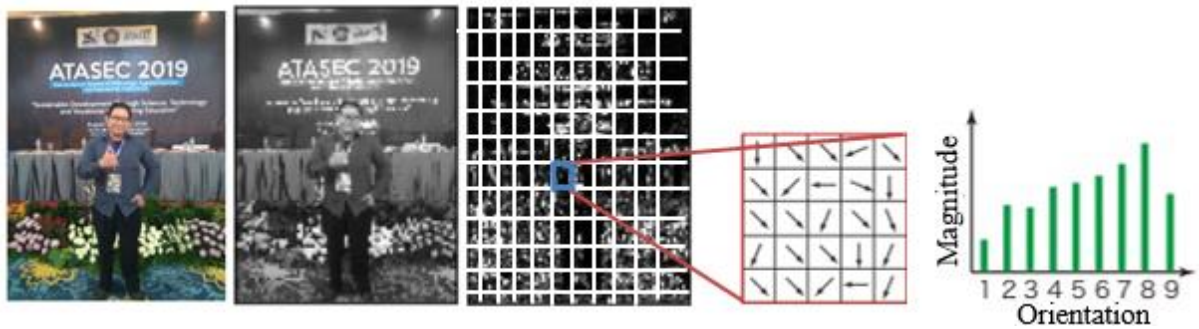

Figure 6. Histogram of oriented gradients feature extraction.

3. Method

In this study, we propose an application to detect the face by implementing Viola-Jones Haar Cascade and compared with the Histogram of Oriented Gradients. The following steps for image processing that have been done in both methods are shown in figure 7 and figure 8.

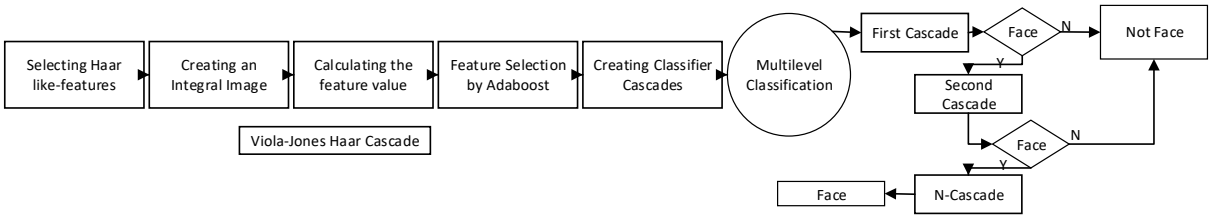

Figure 7. V-J in general.

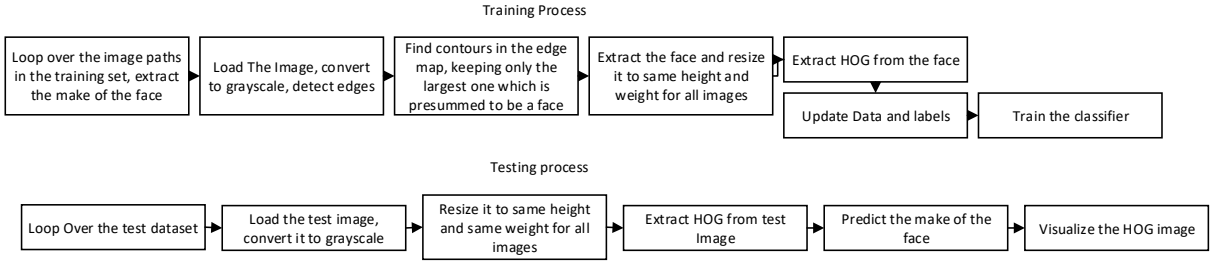

Figure 8. HOG in general.

The entire detailed process of training and testing using HOG can be seen above, in this study the proposed method used for training following: Sampling P (positive samples) from our training data of the face that we want to detect and extract the HOG descriptors from these samples, then we sampling N (negative samples) from a negative training set that doesn't have or contain any face we want to detect, then extracting HOG descriptors from these sample, which number of dataset N more than P, train a linear Support vector Machine (SVM) on our positive and negative sample. The next step is applying hard-negative mining, for each image after the scaling process in our negative training set, sliding window technique is used across the image, each window calculated the HOG descriptor and the classifier applied, if the classifier showing the wrong classification results in a given window as a face, record the feature vector associated with the false-positive result along with the probability of classification. Next step is sorting the false-positive samples that we found during the stage of hard-negative mining by their confidence of probability, re-retraining the classifier using these hard-negative samples by 3-5 steps and the classifier window is trained and it is ready to be used for the test dataset.

The entire detailed process of face detection using V-J can be seen in Figure 6, in this study the following method had the main process: first we created dataset to be trained, first we import the library that we need, and start capturing video or images, then we detect object in video stream or images using Haar Cascade Frontal Face, and initialize sample face image. We proposed the system so it can initialize face id. The looping process including capturing video frames or images, converting frames or images to grayscale, and detect frames of different sizes using multi-scales. Dataset training for each image frame containing face must be cropped into the rectangle with an increment for each sample face image and saving the captured image into dataset folder. For the training dataset, import the library for image processing and matrix calculation, then create a local binary pattern histogram for face recognition, and using prebuilt frontal face training model for face detection. The next main process is getting the images and label data, get all file path, initialize empty face sample, initialize empty id and loop all the file path, get the image and converting into grayscale, using library to process the image into an array, then get the image id. Get the face from training images and loop for each face, append it to their respective id,

add the image to face samples and add the ID to IDs, pass the face array and IDs array, get the faces and IDs, and train the model using faces and IDs then save it into the model trainer.

The last main process for testing is to import the library for image processing and matrix calculations, then create local binary patterns histogram, load the trained model that created before, load the prebuild model for Frontal Face and create a classifier from the prebuilt model, initialize and start the video frame capture. Convert the captured frame into grayscale space color and get all face from the video frame. For each face in video frames, create a rectangle around the face and recognize the face belongs to their respective id and display the video frame with the bounded rectangle.

4. Results and discussion

Classification process for face detection was done by trained 300 images with various conditions including various scaling, various type of occlusion, heavy makeup on the face, various pose, various expressions and various illuminations for face training and 90 images not included in training dataset as testing data. The following results of face detection using VJ and HOG can be seen in table 2 and table 3.

Table 2. Result face detection using V-J.

Condition of Image	Amount of Test Data	No. of Face Recognized	Recognition Ability
1st Scale	15	13	86.67%
2nd Occlusion	15	4	26.67%
3rd Make Up	15	10	66.67%
4th Pose	15	12	80%
5th Expression	15	13	86.67%
6th Illumination	15	12	80%
Average			71.11%

Table 3. Result face detection using HOG.

Condition of Image	Amount of Test Data	No. of Face Recognized	Recognition Ability
1st Scale	15	15	100%
2nd Occlusion	15	4	26.67%
3rd Make Up	15	11	73.33%
4th Pose	15	13	86.67%
5th Expression	15	14	93.33%
6th Illumination	15	14	93.33%
Average			79%

Table 2 shows that face detection through V-J has good accuracy. In first condition with various scale of dataset without normalizing the accuracy is 86.67%, the second condition with various occlusion is 26.67%, the third condition with using heavy makeup on face is 66,67%, the fourth condition with different pose is 80%, the fifth condition with variety of expression is 86.67% and the last condition with illumination is 80%, that makes the average accuracy of all condition was 71.11%. The error that occurred in the face detection that the face was not detected due to various reasons, mostly because of too much tilting and heavy occlusion. V-J has more false-positive than HOGs and the example of false-positive and true negative can be seen in figure 9.

Figure 9. False-positive and true-negative in V-J.

Table 3 shows that face recognition through the HOGs has more accuracy than the V-J feature. In first condition without normalize the scale of dataset, the accuracy value is still perfect 100%, the second condition with various occlusion was 26.67%, the third condition with makeup was 40%, the fourth condition with different pose was 80%, the fifth condition with variety of expression was 93,33% and the last condition with various illumination was 93.33%. The average accuracy of all condition was 79. In group face considerate success if the total face detected is more than 80% of total people presents in the image. Retried the experiment as much as 5 times with different datasets and increased the dataset for training by 30 each time, the calculation result of accuracy is displayed on the graph below:

	Scale	Occlusion	MakeUp	Pose	Expression	Illumination
Fifth Trial	86.67	33.33	73.33	80	93.33	80
Fourth Trial	93.33	26.67	80	80	93.33	93.33
Third Trial	93.33	33.33	66.67	86.67	86.67	86.67
Second Trial	93.33	33.33	73.33	86.67	93.33	86.67
First Trial	86.67	26.67	66.67	80	86.67	80

	Scale	Occlusion	MakeUp	Pose	Expression	Illumination
First Trial	100	33.33	73.33	86.67	93.33	93.33
Second Trial	93.33	33.33	80	93.33	93.33	86.67
Third Trial	100	26.67	73.33	100	86.67	86.67
Fourth Trial	100	46.67	66.67	100	93.33	93.33
Fifth Trial	93.33	26.67	73.33	100	93.33	86.67

Figure 10. Graph of total trial.

5. Conclusion

Based on the experiment before the system can classify and detect the face in many cases and conditions. With six types of condition and five times of trial, obtain the accuracy by 75,33% by using V-J and 80,22% by using HOG. V-J algorithm can detect frontal face very well in images, regarding of their scale, pose, makeup, expression, and illumination, but rather difficult to detect the face who have occlusions like using helm, eyeglass, and mask. The V-J algorithm can perform in real-time on many applications and hardware, the main problem with Haar cascades is in the parameter called detect multiscale and scale factor. If the scale factor is too low, many pyramid layers will be evaluated, this will help to detect more than one faces in images, but the detection process will be slower and increases the false-positive detection rate. On the other hand, if scale factor is too large, it cannot detect the face in small pixel. The recommended size for datasets at least above 250*250 pixels. The HOGs more accurate than V-J for face detection, it can represent local appearance very well.

Acknowledgment
This research is supported by the Department of Electrical Engineering, State Polytechnic of Malang. We also thank for providing facilities in the process of working on it.

References
[1] Simpson E A, Maylott S E, Leonard K, Lazo R J and Jakobsen K V 2019 Face detection in infants and adults: Effects of orientation and color *Journal of experimental child psychology* **186** 17-32
[2] Leo M, Medioni G, Trivedi M, Kanade T and Farinella G M 2017 Computer vision for assistive technologies *Computer Vision and Image Understanding* **154** 1-15
[3] Bradski G and Kaehler A 2008 *Learning OpenCV: Computer vision with the OpenCV library* (O'Reilly Media, Inc.)
[4] Alyushin M V, Alyushin V M and Kolobashkina L V 2018 Optimization of the Data Representation Integrated Form in the Viola-Jones Algorithm for a Person's Face Search *Procedia computer science* **123** 18-23
[5] Nguyen T, Hefenbrock D, Oberg J, Kastner R and Baden S 2013 A software-based dynamic-warp scheduling approach for load-balancing the Viola–Jones face detection algorithm on GPUs *Journal of Parallel and Distributed Computing* **73**(5) 677-685
[6] Zeng D, Zhao F, Ge S and Shen W 2019 Fast cascade face detection with pyramid network Pattern *Recognition Letters* **119** 180-186