ICIC Express Letters
Part B: Applications
Volume **12**, Number **4(tentative)**, April **2021**

ICIC International ©2021 ISSN 2185-2766
pp. 1–ELB-2010-001

FACE RECOGNITION FOR WEARING A VEIL CASE USING HISTOGRAM OF ORIENTED GRADIENTS

Muhammad Arafah[1,2], Andani Achmad[2], Indrabayu[3] and Intan Sari Areni[2]

[1]Informatics Study Program
STMIK AKBA
Jl. Perintis Kemerdekaan KM. 9 No. 75, Makassar, Indonesia
arafah@akba.ac.id

[2]Department of Electrical Engineering
[3]Department of Informatics
Hasanuddin University
Jl. Poros Malino KM. 6, Bontomarannu, Gowa 92119, Sulawesi Selatan, Indonesia
{ andani; indrabayu; intan }@unhas.ac.id

Received October 2020; accepted January 2021

Abstract. *This study aims to identify people based on faces using occlusions in the form of a veil. These results will be used as a reference for tracking suspects in terrorism. The data were collected using Closed-Circuit Television (CCTV) as in the passenger inspection area at the airport. The data used were classified into two parts namely training and testing data. In the training data, the photo data used consists of four angles per person, while the testing data were videos consisting of data with and without occlusion. The feature-based facial identification method used refers to three local features namely the mouth, nose, and a pair of eyes with the Viola-Jones method. The face detection results obtained were processed through pre-processing, feature extraction by the Histogram of Oriented Gradients (HOG) method, and classification by the Multi-class Support Vector Machine (MSVM) method. The level of accuracy with occlusion obtained was 75.08%.*
Keywords: Suspects in terrorism, Face recognition, Occlusion, HOG, Multi-class SVM

1. **Introduction.** In recent years, terrorism has become increasingly crazy and dangerous with brutal acts to frighten the public. Terrorism not only kills innocent people but also greatly affects the development and progress of the national economy in the social sphere [1]. Terrorist attacks in the past have caused enormous damage to a country's life, economy, infrastructure, and its recovery is a very tough task [2]. To avoid criminal acts of terrorism, the Indonesian government takes steps to anticipate continuously based on the principles of human rights protection and prudence. The government issued a regulation on the eradication of criminal acts of terrorism, which was then included in the National Research Master Plan for the Ministry of Higher Education and Technology of the Republic of Indonesia for 2017-2045 [3].

One of the most important public facilities and often targeted in protection efforts is airports, and this can be seen from the frequent simulations of handling terrorism acts, for example, a simulation of, simulations of Jet Air's RP-R0001 airplane demonstration hijacked by six terrorists at Soekarno-Hatta Airport in Tangerang in Terminal 1A. The terrorists crept into the plane area with full fire, several officers of Soekarno Hatta Airport's Aviation Security (AvSec) were taken as hostage. This information was reported in the news reporters of tribunnews [4]. The management of PT Angkasa Pura I (Persero) as The Manager of I Gusti Ngurah Rai International Airport Bali, increased Security and Personnel Vigilance from Terrorism Threats. General Manager of PT Angkasa Pura I

DOI: 10.24507/icicelb.12.04.xxx

(Persero) Branch Office of I Gusti Ngurah Rai International Airport Bali, Haruman Sulaksono said, "As a vital national object that plays an important role in the lives of many people, a safe and comfortable airport situation is important to create" [5].

In recent years, research related to computer vision, pattern recognition and facial recognition techniques has shown promising performance, but in real-life images or videos, various occlusions are often found on the human face, such as sunglasses, masks and wearing a veil [6,7]. This research is a continuation of previous research conducted by the authors with the results of the best distance between the position of the walkthrough metal detector and CCTV in identifying the faces of passengers who will enter the airport waiting room [8]. The identification of terrorists or suspected terrorists is still based on the faces of the perpetrators, which are usually announced in the form of photographs. Therefore, the development of a terrorism detection system can be started based on facial recognition without occlusion or with occlusion such as the use of a veil to camouflage to avoid not being identified.

Guo et al. in 2018 researched the detection of faces with occlusion based on facial physiology. The training and testing data used were photographs, and the system accuracy was 57.3% better than the Adaboost method with 26% accuracy and 46% seetaface accuracy [9]. In image processing, one of the steps that affect accuracy is preprocessing. Indrabayu et al. in 2017 conducted a study on a method of increasing brightness to improve the quality of the input image before the feature extraction stage. The input data used consist of validation data of 300 colored eye images from 4 people using 40×95 pixels. The results of system performance produce an average accuracy of 93.5%. Then in 2019 the authors conducted a study to determine the level of accuracy in the detection of eye, nose and mouth features. They use the Viola-Jones method to detect faces using a data input system in the form of video data [10,11].

In 2019, Matsumura and Hanazawa proposed a method for human detection using color contrast-based HOG, in which a gradient-oriented histogram is created by calculating the gradient based on the similarity of colors in the local area or cells of an image. Neighbor cells will be captured by shifting one pixel upward, left, up and down. Furthermore, it will use real Adaboost, support vector machine and random forest for training. The proposed experiment is comparing the classification performance of the proposed method with HOG and comparing the detection performance of the proposed method with HOG. Then the evaluation methods used in this experiment are the False Positive per Window (FPPW), the False Positive per Image (FPPI), and the error tradeoff detection curve. The dataset used is the INRIA Person and the NICTA Pedestrian datasets. The results showed that the INRIA Person dataset using real Adaboost based classification reduced the error rate by 1.8%, while the SVM classifier reduced the error rate by 25.2% and random forest classification reduced the error rate by 23.7%. Meanwhile, using the NICTA Pedestrian dataset, classification based on real Adaboost reduces errors by 2.7%, then SVM-based classifiers reduce errors by 11.8% and classification based on random forest reduces errors by 17.8%. The focus of this study is the detection of human presence using color contrast-based histograms of oriented gradients. Meanwhile our research is focused on identifying a person by referring to the face area [12].

Based on those researches, it was conducted using Viola-Jones for face detection, and then a pre-processing process was carried out to improve image quality. Meanwhile, for the feature extraction, it used HOG method. While the MSVM method was used for the classification stage, both for face detection systems without using occlusion or with occlusion. This research is a continuation of previous research conducted by the authors with the results in the form of the best distance between a position of a CCTV and a walk through the metal detector in identifying the faces of passengers who will enter the airport waiting room. The best distance obtained was 300 cm [8]. This distance is used in testing data collection in this study.

The organization of this paper is as follows. In Section 1 the background of the paper is introduced. The proposed system is explained in Section 2. The results are discussed in Section 3 and finally, Section 4 concludes the paper and discusses the future work.

2. **Proposed Method.** The system design used generally consists of two parts, namely the training and testing stage as shown in Figure 1.

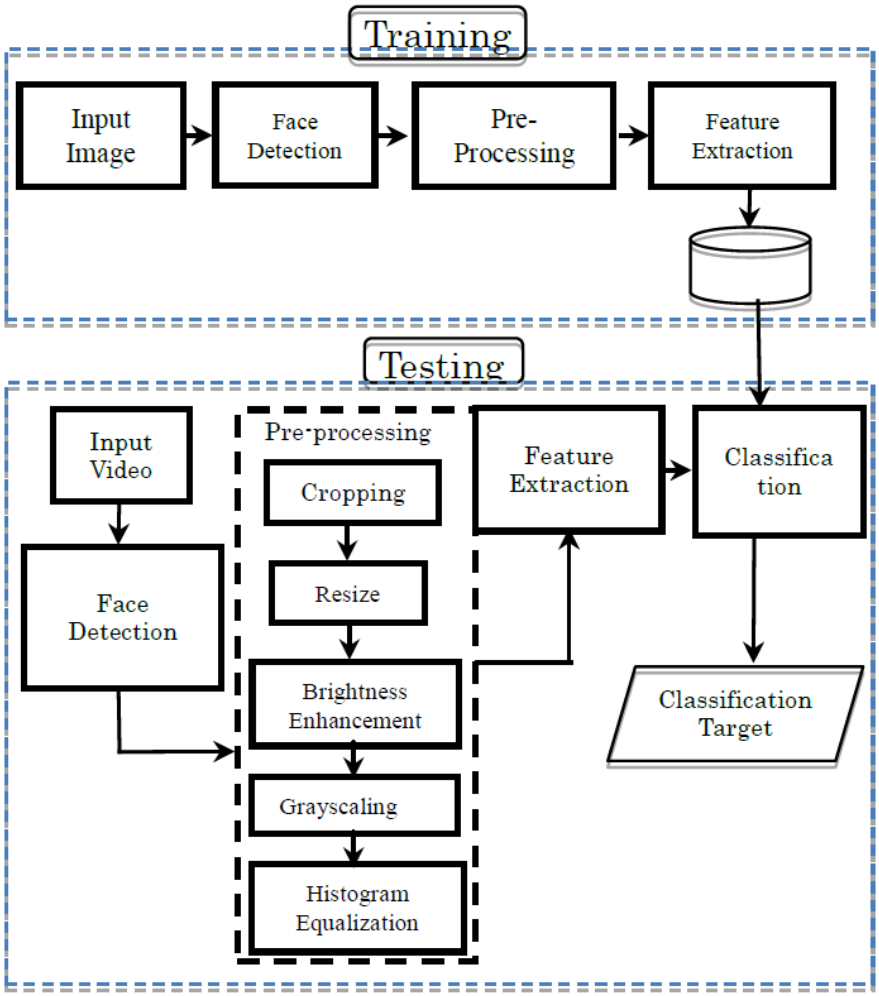

FIGURE 1. System design

2.1. **Input data.** Input data were classified into data for the training process and input data for the testing process. Training data in the form of image data (photo) taken used a single-lens reflex digital camera (DSLR) with a camera distance from the object about 50 cm. Illustration of training data retrieval is shown in Figure 2. The training data consists of 9 targets in the form of photo with four different angles as shown in Figure 2. There are two classifications of obtaining the testing data, namely, testing data with and without occlusion. The occlusion in this paper uses a veil. The medium used for the data testing was CCTV where the position of the CCTV was placed at the height of 250 cm from the floor. The distance between CCTV and a walk through metal detector device is 300 cm as shown in Figure 3. For the training process, input data consists of nine face photos where each face (F) of the target has four different angles (A). For the first angle (A-1) facing forward vertically, while for the second angle (A-2) facing downward with a shift of about 15 degrees from position (A-1). For the third angle (A-3), get a shift of 15 degrees to the right from the position of (A-1), and then for the fourth angle (A-4), get a shift towards the left about 15 degrees from the position of (A-1) as shown in Figure 2.

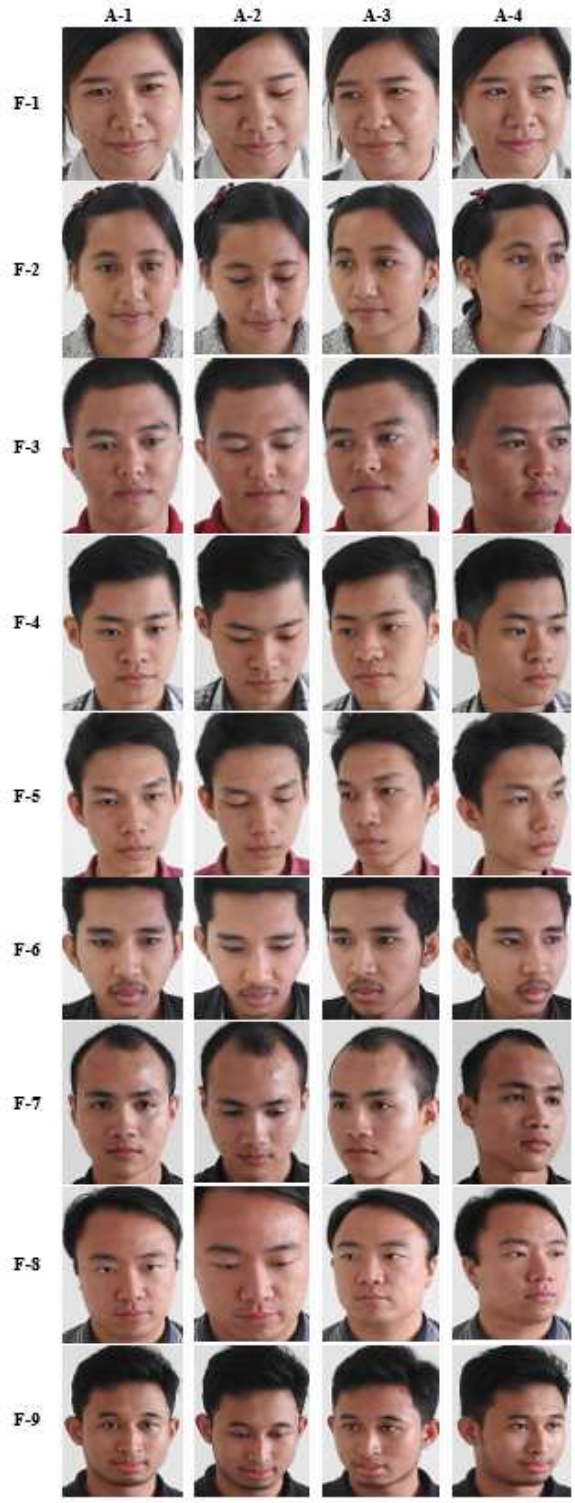

FIGURE 2. Training data

At the stage of the testing data input, there are two classifications of video input data for each object, namely video data with and without veil occlusion. The next step will be the process of acquiring a video frame in the form of a frame and will be processed by the Viola-Jones method to detect local features, i.e., mouth, nose, and eyes.

2.2. **Face detection with Viola-Jones.** Viola-Jones uses four keys in detecting facial features, namely, Haar-like features, integral image, boost learning, and cascade classifier [13].

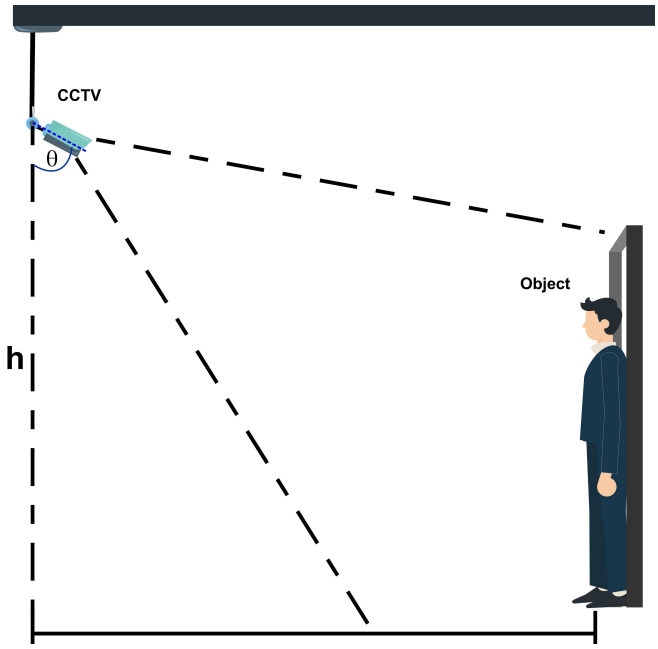

FIGURE 3. Illustration of obtaining the testing data

2.2.1. *Haar-like features.* The Haar-like feature value can be obtained from the difference in the number of bright area pixel values with the number of dark area pixel values, as shown in Equation (1).

$$F_H = \sum F_{White} - \sum F_{Black} \tag{1}$$

where F_H is Haar-like feature, $\sum F_{White}$ is pixel value in bright areas and $\sum F_{Black}$ is pixel value in dark areas.

2.2.2. *Integral image.* Integral image is a process of calculating integral values by referring to Haar-like features that are formed to make it easier to get the difference between dark and light areas.

$$ii(x,y) = \sum x' \leq x,\ y' \leq y^{i(x',y')} \tag{2}$$

where $ii(x,y)$ is integral image and $i(x,y)$ is original image with condition:

$$ii(x,y) = i(x,y) + ii(x-1,y) + ii(x,y-1) - ii(x-1,y-1) \tag{3}$$

2.2.3. *Machine learning Adaboost.* Machine learning Adaboost is a learning phase using the Adaboost method by distinguishing strong and weak classifications. This stage will continue to repeat until the classification is getting stronger. At this stage, the Haar-like feature will be calculated in value. If the value obtained is greater than the predetermined threshold value, then the Haar feature is classified in Adaboost learning. The process will continue to repeat until all Haar features have been used. To determine a weak classifier can be seen in the following equation.

$$h_j(m) = \begin{cases} 1, & \text{if } p_j f_j(m) < p_j \theta_j(m) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $h_j(m)$ is weak classifier, p_j is parity to j, θ_j is threshold to j and m is sub image dimensions.

2.2.4. *Cascade classifier.* Cascade classifier is a stage to classify each classification of the Adaaboost process in determining faces or not faces. In this process, several levels are used to ascertain the face or not the face. One of the characteristics of Viola-Jones is the cascade classifier. This classification algorithm consists of several levels. Each level produces a sub-image that is not a face because it is easier to evaluate than a sub-image

containing a face. Each sub-window will be compared with every feature in each stage. If it does not reach the target, the sub-window will move to the next sub-window and will do the same calculation as the previous process. By referring to the first level classification, each sub-image in the sub-window is classified using several Haar-like features. If the sub-image reaches a threshold, the process will continue to the next stage. However, if it does not reach the threshold, the sub-windows will be rejected, and the process continues to the next sub-image. In the next process, the results are obtained, i.e., sub-windows are detected as faces and continue to the next sub-image. The process will be repeated until finally obtained strong candidates identified as a face [11].

2.3. **The pre-processing stage.** Pre-processing is a step that is done to improve the quality of the image. The image is obtained from the face detection using the Viola-Jones method. Pre-processing used in this study consists of five parts, namely, cropping, resizing, brightness enhancement, grayscaling, and histogram equalization. The explanation of the stages is described as follows.

2.3.1. *Cropping.* Cropping is a process that is carried out to create a new image that is sourced from existing images. The cropping process is carried out by referring to the value of each side of the face bounding box that has been formed, as shown in Figure 4.

FIGURE 4. Cropping on the poses pre-processing

2.3.2. *Resizing.* Resizing is a process carried out by referring to the features detected using the Viola-Jones method, namely by changing the image to equalize the input dimensions of the system. The specifications of the system input data are as follows: the size for face features is 130×110 pixels, while the size for mouth features is 30×50 pixels, for the nose features is 35×40 pixels while for the eye pair features is 20×65 pixels.

2.3.3. *Brightness enhancement.* Brightness enhancement is a process carried out to improve the quality of lighting in the image by adding contrast values to each pixel of RGB (Red, Green, Blue). The following is an example of the brightness enhancement output shown in Figure 5 in point (b).

2.3.4. *Grayscaling.* Grayscaling is a process carried out to convert RGB images into grayscale images. The results of this image conversion can be seen in Figure 5 in point (c).

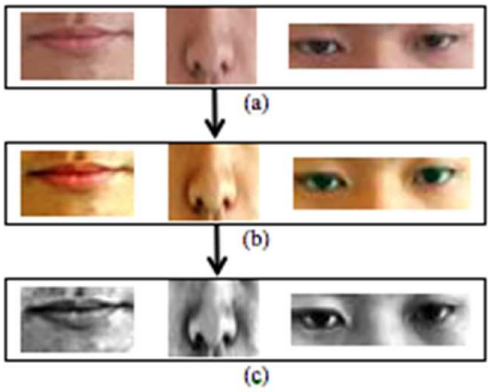

FIGURE 5. (color online) (a) Input image, (b) output of brightness enhancement, and (c) output of grayscaling

2.3.5. *Histogram equalization.* At the histogram equalization stage, the method used in this process is CLAHE (Contrast Limited Adaptive Histogram Equalization), and this is done for histogram leveling of face images that have passed the grayscaling process. The CLAHE method works by limiting the specified contrast level (clip limit) of the image to avoid excessive contrast. The following is the determination of the CLAHE parameter used, while the value of the clip limit parameter is 0.005 and NumTile with size [2 2].

2.4. **The feature extraction stage.** The use of extraction feature with HOG is done after the pre-processing stage. The use of extraction feature with HOG is done after the pre-processing stage based on cell size parameters [2 2]. The stages of this method can be seen in Figure 6. Some examples of feature patterns are the mouth pattern with a size of 30×50 pixels as shown in Figure 7(a), pattern of nose features with size of 35×40 pixels as shown in Figure 7(b) and feature pattern of a pair of eyes with a size 20×65 pixels, which can be seen in Figure 7(c).

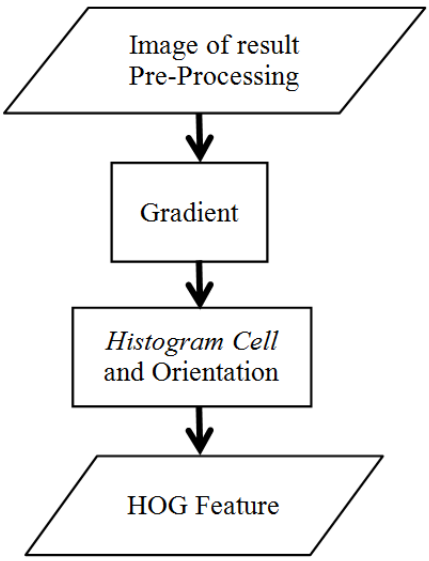

FIGURE 6. Stage of HOG method

2.5. **The classification stage.** The use of classification from the extraction of histogram of oriented gradients features uses the MSVM method. The classification process is divided into two parts, namely classification in the training process and classification in the testing process. In the training process the system will classify 9 faces, where each face has 4

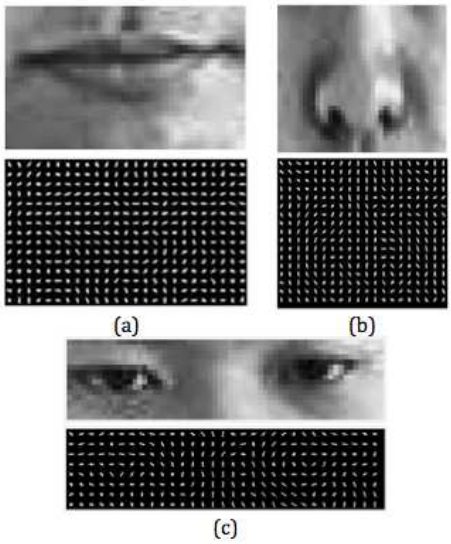

FIGURE 7. The example of feature pattern: (a) mouth, (b) nose, and (c) a couple of eyes

angles, so the class that will be formed is "class = [1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5 6 6 6 6 7 7 7 7 8 8 8 8 9 9 9 9]". This feature will be classified in the form of vector $m \times r$, where m is the number of faces to be trained while r is the value of features that will be owned by the face, namely the value of face features in the form of mouth, nose and a pair of eyes. The next step is to create a train feature model that functions to match the train features with the available classes. The train feature model will produce a 1×1 struct file. The training model that has been created using the predict function on the same feature will be tested with an accuracy of up to 100%.

Furthermore, in the testing process, faces in the form of video frames will be classified according to the class provided, i.e., "class = [1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5 6 6 6 6 7 7 7 7 8 8 8 8 9 9 9 9]". The new features tested are matched with the trained model that has been created. The MSVM method will classify the value of incoming features, in accordance with the value of the features in the trained model. Features will be classified in the form of $n \times p$ vectors, where n is the number of faces in a video frame that has a complete feature cut (full face), while p is the value of the features the face has, namely the face feature value in the form of a mouth, nose and a pair of eyes.

3. **Result and Discussion.** Face identification or face recognition, either using occlusion or without using occlusion, shows the results as in Figure 8. This result is obtained by ignoring the classification of local face features, namely the mouth, nose, and a pair of eyes that have been processed using HOG feature extraction. Based on Figure 8, face

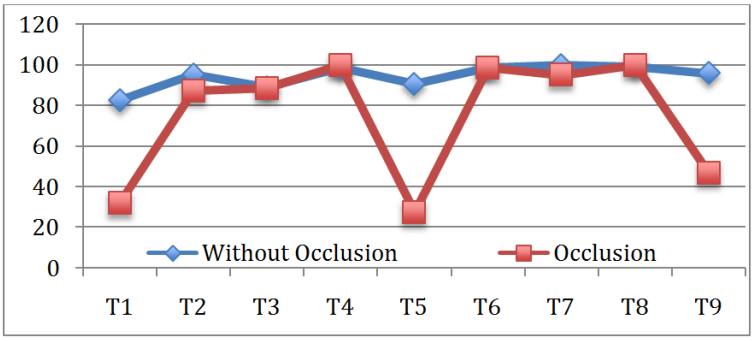

FIGURE 8. Comparison chart of face recognition with and without occlusion

recognition using the HOG and MSVM classification for faces without occlusion has an accuracy of 94.46%, while for face recognition with a veil occlusion the accuracy is 75.08%. The comparison of face recognition with and without occlusion can be seen in Figure 8. There were three targets in the figure which have different face recognition gaps, that is, T1 target, T5 target and T9 target. In the T1 and T5 targets for face recognition using occlusion, the level of accuracy was extremely low in which many frames in T1 and T5 were predicted as T8, after that many frames in T9 target is predicted as T2.

4. **Conclusion.** This research has succeeded in identifying face with veil occlusion or without occlusion by using the HOG feature extraction method and classification methods with the MSVM. In this study, two data classifications are used, namely, training data and testing data. In the training data using nine photos with four different angles, while for testing data using video data taken using CCTV with two conditions, namely, data using occlusion in the form of a veil and data without using occlusion. The results of the study showed that the overall accuracy of the face recognition system for conditions using a veil occlusion was 75.08%, while the overall system accuracy without using occlusion was 94.46%. In the future, we will find the best technique for face identification using the occlusion of eyeglasses, a mustache, or beard.

Acknowledgment. The authors would like to thank for the financial support of this work with a Doctoral Dissertation Research grant from the Ministry of Research Technology and Higher Education Indonesia and LPPM – Hasanuddin University (UNHAS) 2019.

REFERENCES

[1] X. Huang, H. Zhang, H. Yin and X. Z. Gao, Dynamic study of a counter-terrorism model, *2016 the 35th Chinese Control Conference (CCC)*, pp.10328-10332, doi: 10.1109/ChiCC.2016.7554990, 2016.

[2] S. Singh, D. Indurkhya and A. Tiwari, An avant-garde approach for detection of key individuals with leader hierarchy determination using FIMAX model (anti-terrorism approach), *2018 Int. Conf. Inf. Manag. Process. (ICIMP2018)*, pp.89-99, doi: 10.1109/ICIMP1.2018.8325847, 2018.

[3] T. The ministry of Research, *National Research 2017-2045 (28 February 28, 2017)*, http://simlitabmas.ristekdikti.go.id/, vol.2045, p.96, 2017.

[4] M. Pt et al., *Anticipation of Terror, Ngurah Rai Airport Increases Security*, https://nasional.republika.co.id, 2019.

[5] Z. A. Setiawan and H. Tangerang, *Terrorists, Plow Plane in Soekarno Hatta Airport Simulation, Using Weapons and Threatening to Blast Aircraft*, https://wartakota.tribunnews.com, 2019.

[6] F. Zhao, J. Feng, J. Zhao, W. Yang and S. Yan, Robust LSTM-autoencoders for face de-occlusion in the wild, *IEEE Trans. Image Process.*, vol.27, no.2, pp.778-790, doi: 10.1109/TIP.2017.2771408, 2018.

[7] F. Cen and G. Wang, Dictionary representation of deep features for occlusion-robust face recognition, *IEEE Access*, vol.7, pp.26595-26605, doi: 10.1109/ACCESS.2019.2901376, 2019.

[8] M. Arafah, A. Achmad, Indrabayu and I. S. Areni, Face recognition system using Viola Jones, histograms of oriented gradients and multi-class support vector machine, *J. Phys. Conf. Ser.*, vol.1341, no.4, doi: 10.1088/1742-6596/1341/4/042005, 2019.

[9] Z. Guo, W. Zhou, L. Xiao, X. Hu, Z. Zhang and Z. Hong, Occlusion face detection technology based on facial physiology, *Proc. of the 14th Int. Conf. Comput. Intell. Secur. (CIS2018)*, pp.106-109, doi: 10.1109/CIS2018.2018.00031, 2018.

[10] Indrabayu, R. A. Tacok and I. S. Areni, Modification on brightness enhancement for simple thresholding in eyelid area measurement, *ACM Int. Conf. Proceeding Ser.*, pp.101-104, doi: 10.1145/3121138.3121197, 2017.

[11] Indrabayu, Nurzaenab and I. Nurtanio, An approach in auto valuing for optimal threshold of Viola Jones, *J. Phys. Conf. Ser.*, vol.1198, no.9, doi: 10.1088/1742-6596/1198/9/092003, 2019.

[12] R. Matsumura and A. Hanazawa, Human detection using color contrast-based histograms of oriented gradients, *International Journal of Innovative Computing, Information and Control*, vol.15, no.4, pp.1211-1222, doi: 10.24507/ijicic.15.04.1211, 2019.

[13] P. Viola and M. Jones, Managing work role performance: Challenges for twenty-first century organizations and their employees, *Rapid Object Detect. Using a Boost. Cascade Simple Featur.*, pp.511-518, 2001.