Research Article

Adapting Viola-Jones Method for Online Hand / Glove Identification

Taib Shamsadin Abdulsamad[1,2,a*], **Mahmud Abdulla Mohammad**[1,3,b], **Faraidoon Hassan Ahmad**[4,5,c]

[1] Department of Computer, College of Basic Education, University of Raparin, Sulaymaniyah, Iraq
[2] Department of Computer Science, College of Science and Technology, University of Human Development, Sulaymaniyah, Iraq
[3] Department of Information Technology, College of Science and Technology, University of Human Development, Sulaymaniyah, Iraq
[4] Department of Pharmacognosy & Pharmaceutical Chemistry, College of Pharmacy, University of Sulaimani, Sulaymaniyah, Iraq
[5] Department of Information Technology, University College of Goizha, Sulaymaniyah, Iraq
E-mail: [a]taib.shamsadin@uor.edu.krd, [b]MohammadMA@uor.edu.krd, [c]faraidoon.ahmad@univsul.edu.iq, [c]faraidoon.ahmad@uog.edu.iq

Access this article online		
Received on: December 30, 2020	**Accepted on:** Febraury 25, 2021	**Published on:** June 30, 2021
DOI: 10.25079/ukhjse.v5n1y2021.pp80-90	**E-ISSN:** 2520-7792	

Abstract

This article proposes a method for hand identification, adapting the method of Viola-Jones for identifying two different objects. The main objective of this work is to solve the problems of hand identification. Thus, our approach based on learning for two objects as one package. Also, the proposed method folds into three parts; the first part is training for both objects, second detection of both objects, and third the identification step to identify if the hand is wearing a glove or not, then labeling each one with a suitable state. Moreover, to test our method, we have proposed a new dataset, which includes a variety of cases with different compositions of hand. As a result, 8 cases were used to test the method. The method was able to detect a human hand successfully. Additionally, it could identify whether the hand was or was not wearhing a glove. The accuracy of detecting a hand without a glove was about 63%, and the accuracy of detecting a hand with a glove on was about 61%. Even though the tests scored different accuracy, as a first step towards solving this problem, it is a big achievement to even reach this level of accuracy.

Keywords: Computer Vision, Image Processing, Object Detection, Viola-Jones, Hand Detection, Identification.

1. Introduction

Nowadays, due to the fast growing use of image processing, it forms a core research area within engineering and computer science disciplines. Digital image processing techniques help in the manipulation of the digital images by using computers. Image processing has numerous applications like visual inspection, remotely sensed image analysis, medical diagnosis, defense surveillance, content-based image retrieval (CBIR), image and video compression, moving object tracking etc. (Acharya & Ray, 2005).

An object detector's objective is to find or recognize all object instances of one or more given object class regardless of scale, location, pose, view with respect to the camera, partial occlusions, and illumination conditions (Verschae & Ruiz-del-Solar, 2015). Object detection has been playing a key role in many applications, which arise in many different fields including industrial automation, consumer electronics, medical imaging, military, video surveillance (Murthy et al., 2020), food safety (Cevallos et al., 2020), autonomous vehicles, and situational awareness (Mohammad et al., 2016;

Muhammad, 2016). More precisely, the applications of object detection include; pedestrian detection, road detection, lane detection, obstacle detection, face detection, crop detection, and hand detection.

Hand identification is considered as an important application that is strongly connected to our health. Indeed, in some circumstances, it is crucial to monitor people for checking to ensure whether they are wearing gloves or not especially, in industrial related, food related, and patient related environments. Also, according to WHO reports, wearing gloves and a mask are two important factors to reduce the transmission of COVID-19 pandemic disease (Ahmed et al., 2020; Dey et al., 2021). Thus, in these situation, especially in medical centers, monitoring people to ensure that they are wearing gloves is crucial. Hand identification methods offer a monitoring process by identifying the hands of those people who are not wearing gloves, along with those who are wearing them.

Generally, hand detection, is the process of extracting and bounding a box of the hand region from a given scene. It is an advanced topic and has received more attention from researchers for hand gesturing and posturing recognition systems.

A number of detection methods have been used in the literature, still Viola-Jones (Viola & Jones, 2001) is one of the fastest and more robust learning-based object detector methods with high detection rate, and it plays an important role in many detection and recognition fields. The Viola-Jones is a well-known and robust appearance-based face detector method. Firstly, the query image is represented in the form of an "Integral Image", which makes feature computation very fast, the integral image for any pixel is equal to the sum of pixels above and to the left of it. Viola-Jones uses AdaBoost classifier that interactively builds a powerful classifier from a conjunction of simple classifiers with specific weights, a series of simple classifiers applied to every sub-region in the image, the sub-region classified as "Not Face" if it fails to pass in any classifier. When a classifier passes an image region, it goes to the next classifier in the series, the image region will be classified as "Face" if it passes through all classifiers in the series (Hendra et al., 2019).

Authors (Da'San et al., 2015; Hazim et al., 2016) used the Viola-Jones algorithm for face region detecting and cropping for face recognition systems. Ahmad (2015) presented a real time ethnicity identification system which the Viola-Jones method applied to extract the face area from the rest of the images. Mathias and Matthew (Kolsch & Turk, 2004), proposed a detection method depending on Viola-Jones with three contributions: frequency analysis-based method for instantaneous estimation of class separability without the need for any training. They built detectors for the most promising candidates and they discovered that with more expressive feature types the classification accuracy increases.

In Nguyen et al. (2012), based on Viola-Jones' work a new approach was addressed for hand detection by detecting the internal region of the hand using its local features without a background. Chouvatut et al. (2015) solved the problem of hand detection from various orientation angles of hand positions using the Viola-Jones detector and SAMME classifier. An automatic hand gesture recognition framework was prevented using the steps in the Viola-Jones method for detection and for the recognition phase Hu invariant moments feature vectors of the detected hand gesture are extracted and a Support Vector Machines (SVMs) classifier is trained for final recognition (Yun & Peng, 2009). Kovalenko et al. (2014), proposed a real time system for hand gesture recognition based on the Viola-Jones detector for the hand detection and thereafter used the Continuously Adaptive Mean Shift Algorithm (CAMShift) to track the position of the extracted hand in the image. Mao et al. (2009) combined Viola-Jones' detection algorithm with the skin-color detection method to perform hand detection and tracking against complex backgrounds. The salience and the fast spread of Covid-19 coronavirus epidemic caught the attention of researchers to new research fields. Wang et al. (2020) proposed a system for a facial mask detection task and a masked face recognition task using three types of masked face datasets, including Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD) and Simulated Masked Face Recognition Dataset (SMFRD).

In the literature, much work has been conducted in the area of hand detection, hand identification problems still remain unsolved, which is a difficult topic, due to the fact that a hand with a glove on is very similar to a hand with a glove off. Hence, this work is aimed to propose the Viola-Jones method to be used for hand identification.

To assess the performance of the proposed method, the paper introduces a new dataset consisting of real-world videos illustrating several cases of hands with gloves on and hands with gloves off. The experimental results indicate that the proposed framework is capable of identifying hands with convenience accuracy. The remainder of the paper is organized as follows. In Section 2, the proposed method is discussed. The definition and format of the proposed dataset is discussed in Section 3. Experimental results are given in Section 4. Finally, conclusions and future work are discussed in Section 5.

2. The Proposed Method

The research focuses on identification process by adapting the Viola-Jones algorithm for hand state. It is obvious that the Viola-Jones algorithm has been designed for single object detection. In this work we adapted this method to use to identify two different objects. Thus, our approach was based on learning both objects as one package, i.e., hands with gloves on and hands with gloves off. The method was successfully able to detect a human hand, and additionally identified it with or without a glove.

The proposed method consists of three parts; the first part is training for both objects, the second is detection of both objects, and the third part is the identification step for identifying if a hand is wearing a glove or not and then labeling each one with the suitable state (i.e. the hand with or without a glove). Figure 1 shows the general scheme of the system methodology.

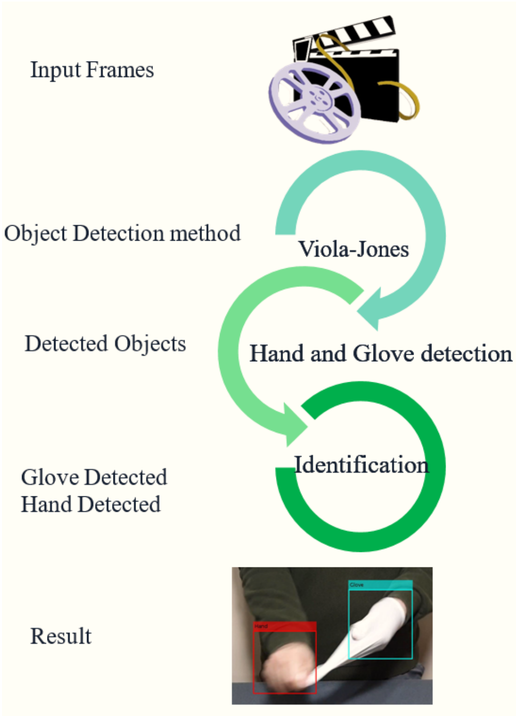

Figure 1. Operation of Proposed Methods.

2.1. Hand detection

The dataset used was prepared for both the training and the testing part. During training for hand detection, the method had to have a positive result and a negative result for the Region of Interest (RoI), thus, a number of images/frames were used to in training to indicate a positive result and a negative result. A positive result showed a cropped hand and a negative result showed no hand at all. Then, these positive and negative results were fed into the Viola-Jones to build a model for detecting a hand during the training step. As a result, an "XML file" was produced and this was known as a model for hand detection. Figure 2 describes and illustrates how to apply the training part for hands without wearing a glove from video frames in our dataset (i.e. Training Data).

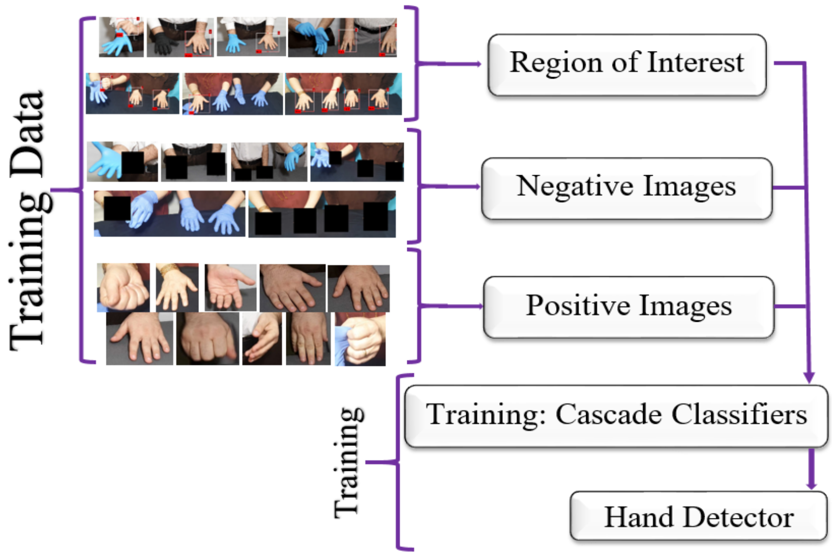

Figure 2. Training steps for hand detection.

This strategy works properly for hand detection from the used data set (i.e. Testing Data). Region of interest (RoI) is a specific zone that identifies pixels of hands inside an image that is extracted from the query frame. The RoI is shown in Figure 3, which is made of hands only. To reach the hand region exactly the area must hold one form of hand shape structure based on the compositions of hands with or without fingers, a closed hand similar to a fist and also the left and right side view and top view etc. 243 positive training images were created (samples are shown in Figure 3(a), along with 155 negative training images (samples are shown in Figure 3(b)).

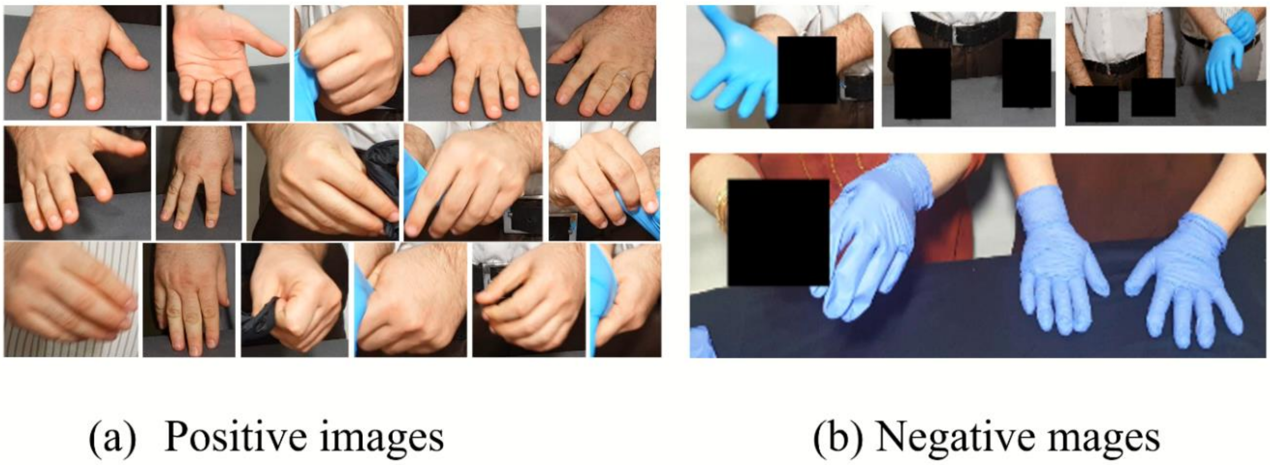

Figure 3. Data Preparation for hand training.

2.2. Glove detection

Preparation for hand with glove on used the same strategy as above. A positive result and a negative result were prepared using images for training and detection. These positive and negative results were fed into Viola-Jones to build a model for detecting a hand with a glove during the training step. As a result, an "XML file" was produced and this was known as the model glove detection. Figure 4 describes and illustrates how to apply the training part for a hand with goves on from video frames in the dataset (i.e. Training Data). In this part, from the training frames, 434 positive training images were created (samples are shown in Figure 5(a)), along with 243 negative training images (samples are shown in Figure 5(b)). The RoI is shown in Figure 5 which shows hands with gloves on only.

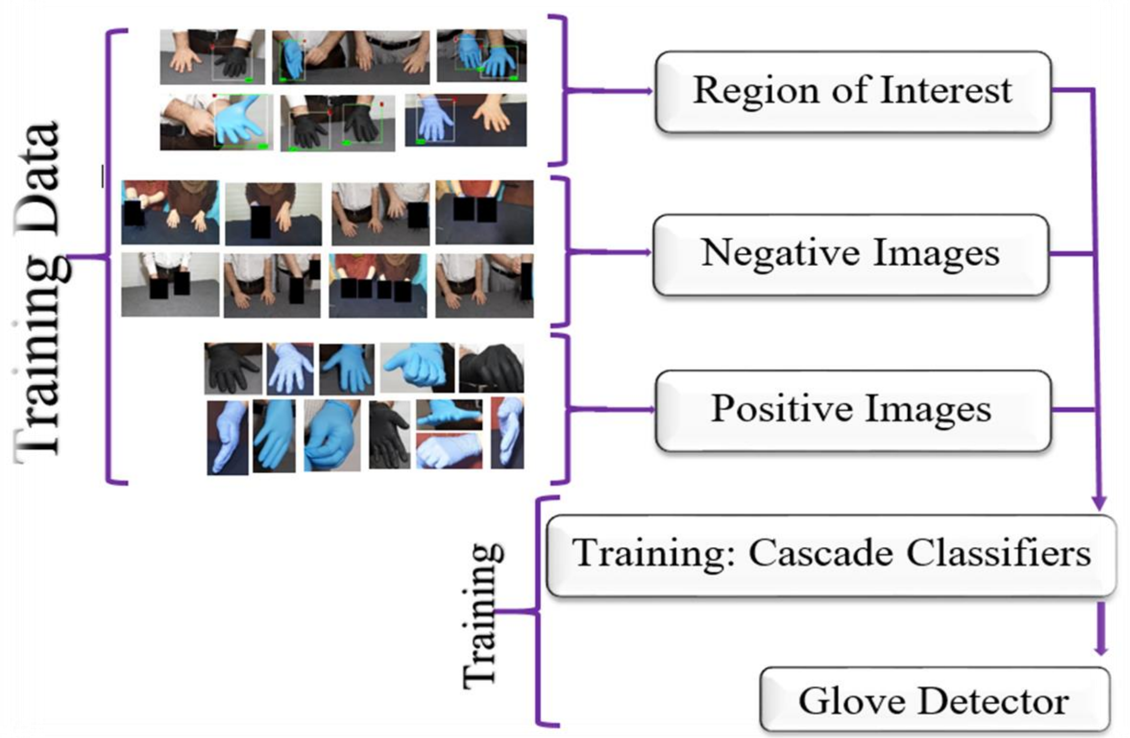

Figure 4. Training steps for glove detection.

Figure 5. Data Preparation for training for hands with gloves on.

2.3. Glove and hand identification

Two models were built and introduced as a result of applying the training steps: one for detecting hands, the other for detecting hands with gloves on. Both detectors were applied as one package - the input frame passes through both detectors. Both RoIs are detected and labeled as GLOVE for hand with a glove or labelled HAND for hand without a glove. Figure 6 illustrates the testing process clearly.

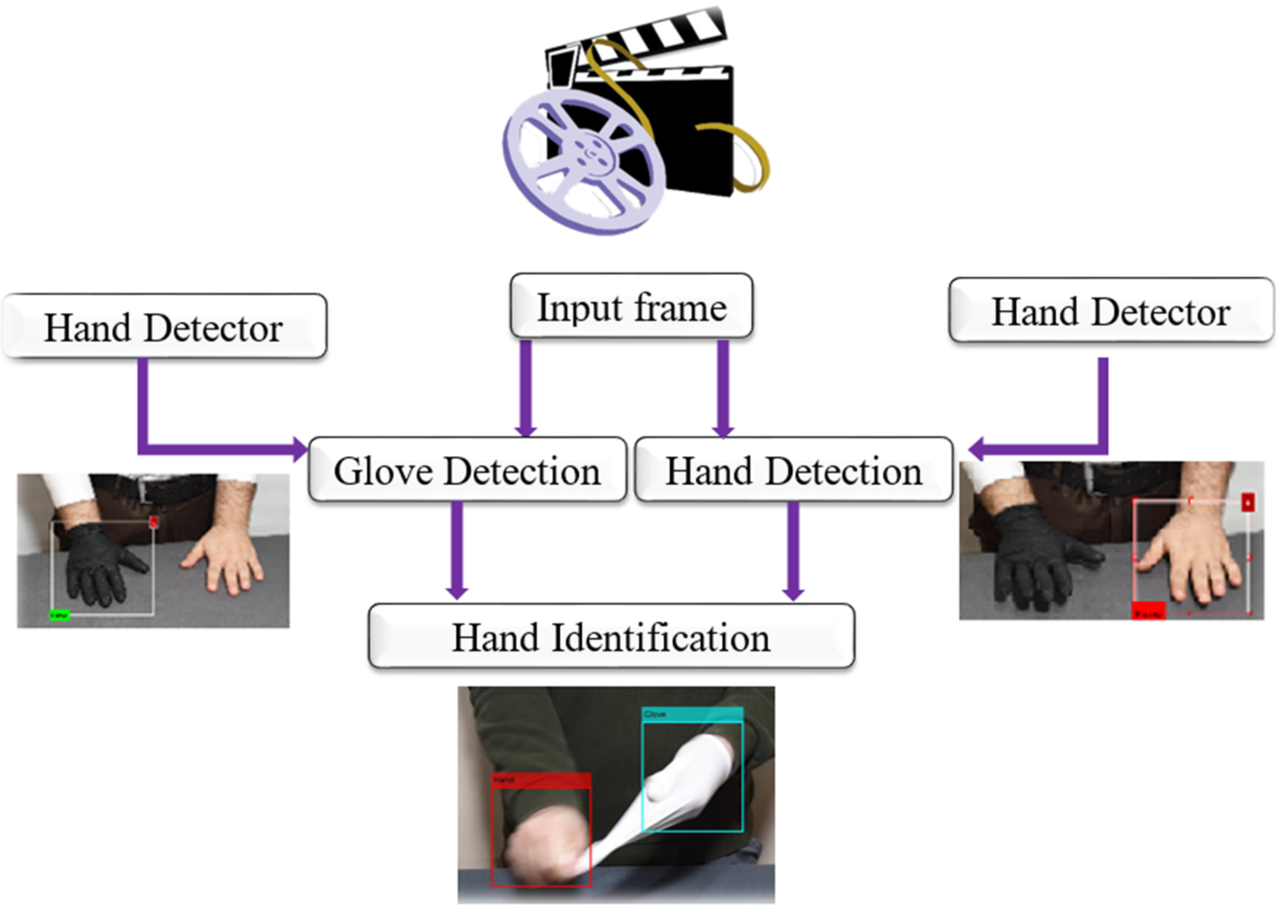

Figure 6. Glove and hand identification methods.

3. Proposed Datasets

The proposed datasets were produced from video frames using specific attributes showing a hand with glove or a hand without a glove. Generally, the dataset contains two main portions: training data and testing data. Training data in the dataset origin was derived from short videos that were recorded under specific measures and decisions that were categorized descriptively in 10 different video sequences under suitable light conditions in addition to uniform backgrounds and using different glove colors. The details of the training dataset are listed in Table 1.

Table 1. Training part of the dataset.

Videos	Descriptions			
	Demonstration		Number of persons	Color
	Hands	Gloves		
1	0	2	1	Black
2	1	1	1	Black
3	2	0	1	No Glove
4	1	1	1	Light Blue
5	0	2	1	Blue
6	3	1	2	Blue
7	0	4	2	Light Blue
8	1	3	2	Light Blue
9	1	1	2	Blue
10	4	0	2	No Gloves

The frame rate of these videos is 30 frame/second. The total frames that were taken from these videos was 2400 frames with "jpg" extension image files. The dimensions of each frame are "3840 * 2160". A number of frames from each situation were collected to make the training dataset. One out of ten frames were chosen because the vidoes usually have consisteny among frames. Consequently, the total frames used for making the training dataset contain 240 frames, which show a variety of different cases.

Likewise, to create the testing part of the dataset in this study, eight different cases were selected. Each case contains 400 frames which show different situations. The total frames used to create the testing dataset are 3200 frames. Tthe details of the cases are provided in Table 2. The dataset is available upon request.

Table 2. Testing part of the dataset.

cases	Descriptions	
	Number of persons	Color
1	2	Blue
2	2	Blue-Black
3	1	Light Blue
4	1	Light Blue
5	1	Blue
6	1	Blue
7	1	white
8	1	White-Black

4. Experimental Results

The proposed method was evaluated on the proposed dataset, which is explained in Section 4. The dataset includes a variety of combinations of images of hands with gloves on and hands with gloves off. The dataset was split into the training and testing subsets. Training frames were used to build the models as explained in Section 3. Also, the testing frames were used to test the method. More precisely, accuracy was calculated for hand identification, results per case, and overall results are reported. The following sections explain the results in detail.

4.1. Hand identification results

In this work, the hands are the region that were focused on. Detected hands have been classified into four classes:

1. True Positive (TP) for hand: means there is a hand in the image and the system detected and recognized it as a hand. This is measured as identifying the hand correctly.
2. False Negative (FN) for hand: means there is a hand in the image and the system detected and recognized it as a hand with a glove on. This is measured as identifying the hand incorrectly.
3. False Positive (FP) for hand: means there is no hand in the image and the system detected and recognized it as a hand. This is measured as identifying the hand incorrectly.

4. True Negative (TN) for hand: means there is no hand in the image and the system does not detect it as a hand. This is measured as identifying that there was no hand correctly.

The accuracy of our system (i.e. accuracy-h) is calculated mathematically using Eq.1. The accuracy equation measures the number of correctly predicted values among the total predicted values of the four hand identification classes.

$$\boldsymbol{Accuracy\_h} = \frac{(\boldsymbol{True\ Positive} + \boldsymbol{True\ Negative})_{\boldsymbol{for\ hand}}}{(\boldsymbol{True\ Positive} + \boldsymbol{True\ Negative} + \boldsymbol{False\ Positive} + \boldsymbol{False\ Negative})_{\boldsymbol{for\ hand}}} \quad \text{…………….. Eq. (1)}$$

Table 3 demonstrates all outcomes of all cases that were produced by the proposed system such as (TP-h, FP-h, FN-h and TN-h). Each case contains 400 frames which were manually calculated for all frames in each case precisely. There are 400 frames per case which shows all classes of hands as shown in Figure 7.

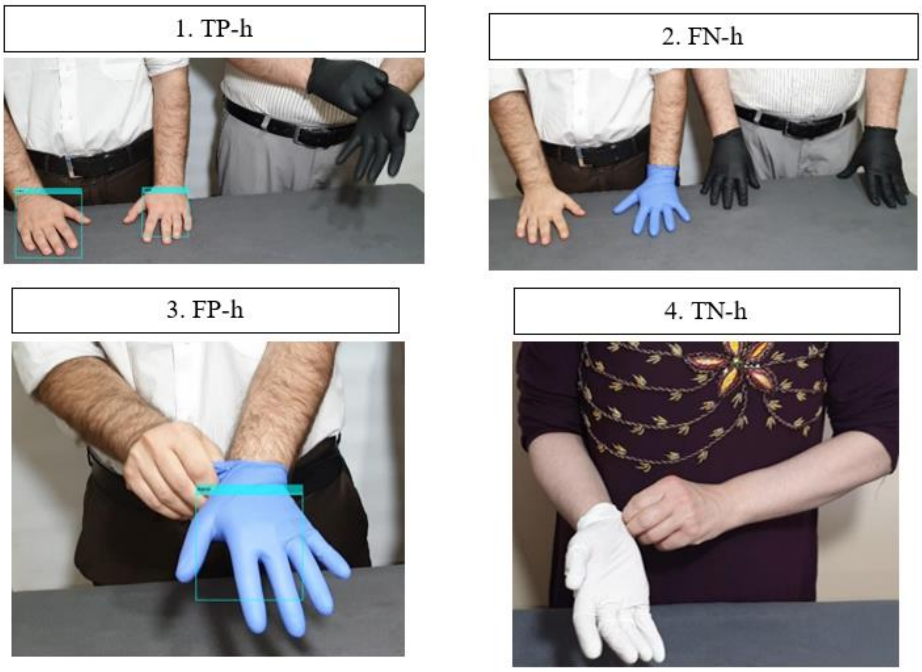

Figure 7. Hand classes.

Table 5 shows the average accuracy of the 8 cases. The experimental results show that the best accuracy reached is (0.787) that is calculated by detecting and labeling objects in each frame [77 True Positive, 85 False Negative, 75 False Positive and 510 True Negative]. The accuracy of detecting each case and overall score of hands with gloves off is shown in Figure 8.

Table 3. Shows the Accuracy of each case and overall score.

	Case1	Case2	Case3	Case4	Case5	Case6	Case7	Case8
TP-h	140	287	171	160	96	105	153	77
FN-h	24	345	214	223	30	272	247	85
FP-h	653	78	50	56	136	166	129	75
TN-h	783	870	365	361	264	257	242	510
Accuracy_h	0.579	0.732	0.67	0.651	0.684	0.452	0.512	0.787

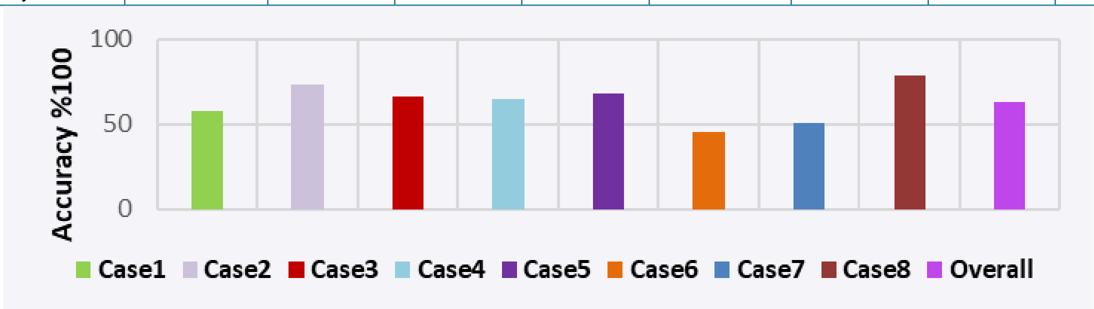

Figure 8. Shows the Accuracy of detecting each case and overall score of hands with gloves off.

4.2. Glove Identification

In this section, the idea is the same as the previous section. The focus is hand identification of hands with gloves on. Four results are possible as follows::

1. True Positive (TP) for glove: means there is a hand with a glove on in the image and the system detected and recognized it as a hand with a glove on. This is measured as identifying the glove correctly.
2. False Negative (FN) for glove: means there is a hand with a glove on in the image and the system detected and recognized it as a hand with a loves off. This is measured as identifying the hand incorrectly.
3. False Positive (FP) for glove: means there is no hand with a glove on in the image and the system detected and recognized it as a hand with a glove on. This is measured as identifying the hand incorrectly.
4. True Negative (TN) for glove: means there is no hand with a glove on in the image and the system did not detect it as a hand with a glove on. This is measured as identifying no hand with a glove on correctly.

The accuracy of this configuration (i.e. accuracy-g) is calculated mathematically using Eq. 2. The accuracy equation simply measures the number of correctly predicted values among the total predicted values of the all 4 hand with gloves on identification classes.

$$Accuracy\_g \; \frac{(True\ Positive + True\ Negative\ )_{for\ Glove}}{(True\ Positive + True\ Negative + False\ Positive + False\ Negative\ )_{for\ Glove}} \quad \text{.........} \ \textbf{Eq. (2)}$$

Table 4 shows all outcomes of all possible statuses that could be produced by the proposed system such as (TP-g, FP-g, FN-g and TN-g). Each case contains 400 frames which were manually calculated for all frames in each case precisely. There are 400 frames per case which show all classes of hands as shown in Figure 9.

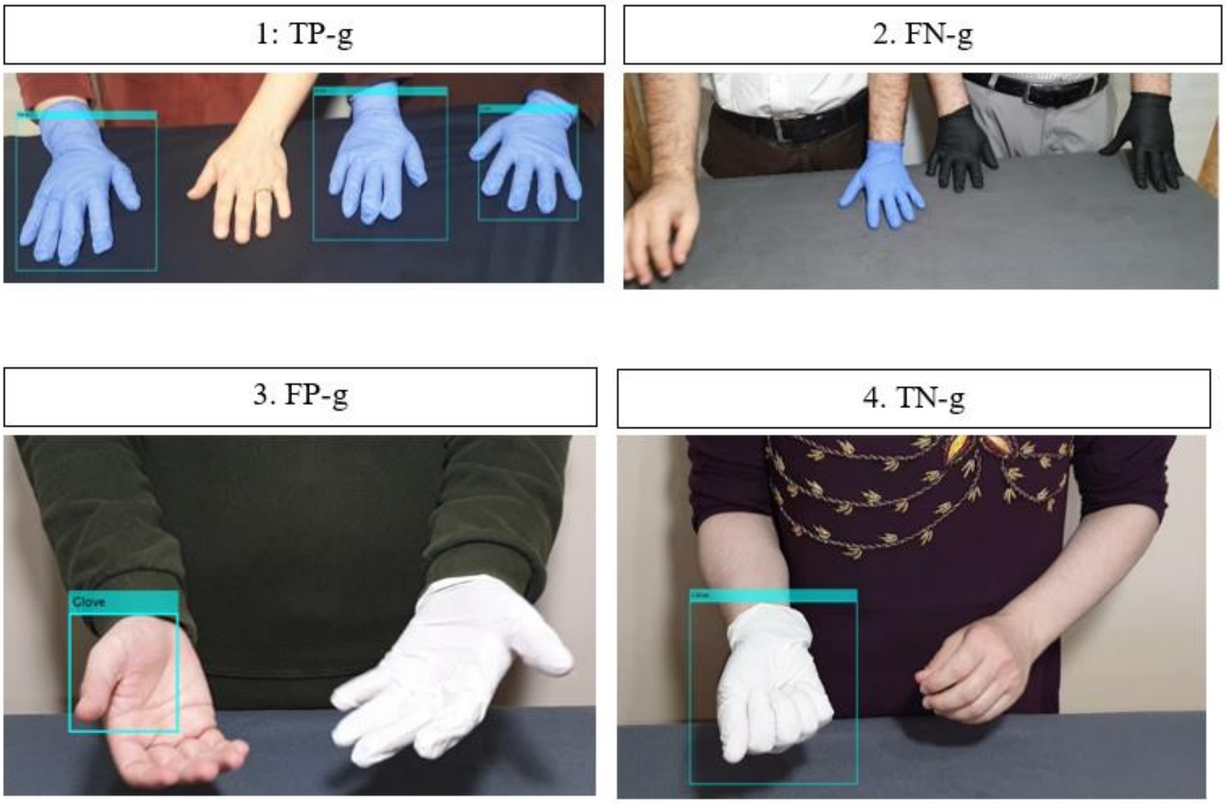

Figure 9. Glove classes.

The accuracy of detection of each case and overall core of hands with gloves on is shown in Figure 10.

Table 4. Shows the accuracy of each case of hands with gloves on.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
TP-g	1840	306	203	195	223	225	213	242
FN-g	196	647	227	226	177	198	158	337
FP-g	91	364	107	101	78	78	92	93
TN-g	77	283	278	286	322	299	308	103
Accuracy-g	0.869	0.368	0.590	0.595	0.681	0.655	0.675	0.445

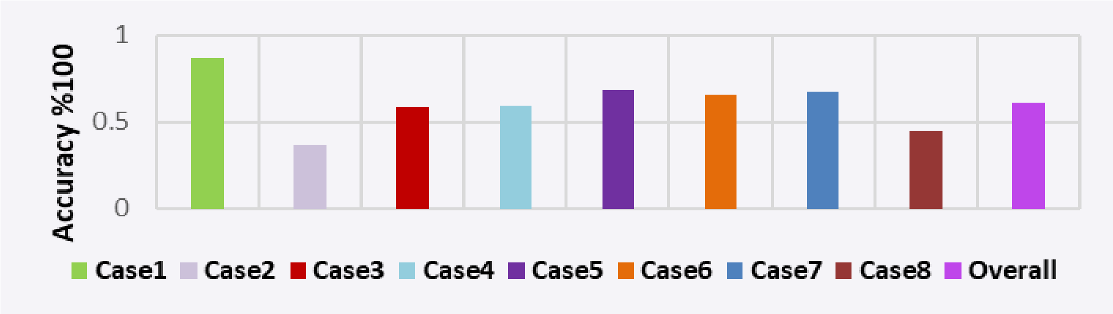

Figure 10. Shows the accuracy of detection of each case and overall score of hands with gloves on.

The experimental results show that the best accuracy reached is (0.869782) that is calculated by detecting and labeling objects in each frame [1840 True Positive, 196 False Negative, 91 False Positive and 77 True Negative].

4.3. Hand/glove identification

As explained in the previous sections, Table 3 shows the accuracy of detecting a hand in each case, then the over all accuracy of detecting a hand with a glove. Table 4, for example, shows the over all accuracy of deteching the hand and the glove and is calculated by taking the average of the accuracy recorded for each case. The accuracy of the proposed method for both objects hand with glove on and hand with glove off, as a first step toward addressing this problem is promising. The accuracy of detecting the hand with a gloves off was about 63% and the accuracy of detecting the hand with glove on was about 61%.

The accuracy of these cases is different from each other, as reported in Table 5 and shown in Figure 11. This can be referred to the diversity of the proposed dataset, which included different colors of gloves and different compositions of hand forms. As the first step toward addressing this problem, it is a big achievement to even reach this level of accuracy.

Table 5. Shows the accuracy of detection of each case for both hand with glove on and hand with glove off.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	AVG
Accuracy_h	0.579	0.732	0.67	0.651	0.684	0.452	0.512	0.787	0.63
Accuracy_g	0.869	0.368	0.590	0.595	0.681	0.655	0.676	0.445	0.610

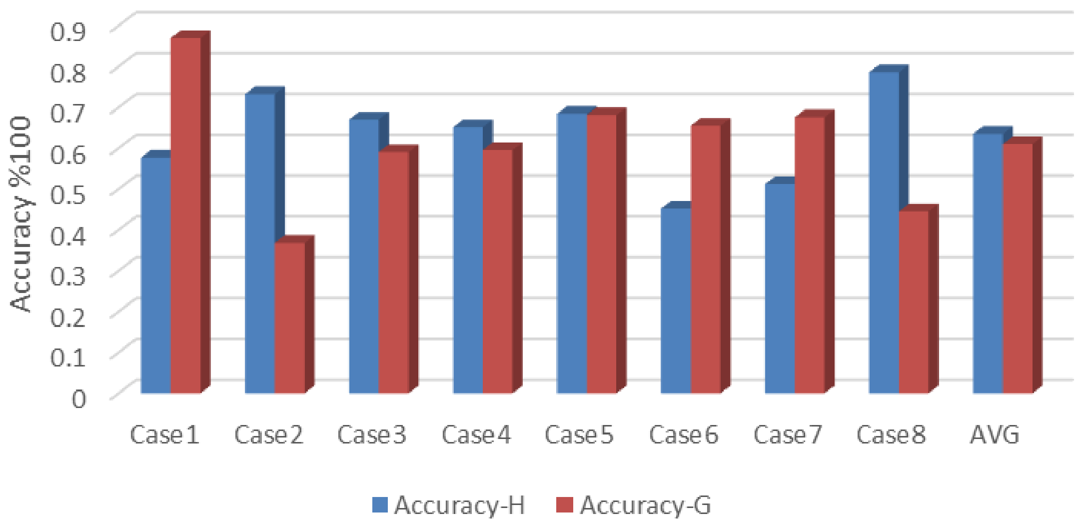

Figure. 11: Shows the accuracy of detecting each case for both hand with glove on and hand with glove off.

5. Conclusion

In this paper a method was proposed for identifying hands with based on adapting the Viola-Jones method for identifying two different objects. The main objective of this work was to address the problem of hand identification in some critical environments. Thus, the approach was based on learning for two objects as one package. Also, the proposed method folds into three parts, the first part was training for both objects, the second was detection of both objects, and the third part was the identification and the labeling of each one with a suitable state.

To test the method, we have proposed a new dataset, which includes a variety of cases with different compositions of hand. Consequently, 8 cases were made inorder to test the method. The method was successfully able to detect a human hand and additionally was able to identify if the hand had a glove on or not. The accuracy of detecting a hand with glove off was about 63%, and the accuracy of detecting a hand with a glove on wasa bout 61%. Although, the cases scored different accuracy, it is refered to the diversity of the proposed dataset, which included different colors of gloves and different compositions of hand forms. As the first step towards addressing this problem, it is a big achievement to even reach this level of accuracy. Of course, there is room to improve the accuracy. Future work should use Random Forest Classifier or Convolutional Neural Network for Detection to explore futher.

References

Acharya, T., & Ray, A. K. (2005). *Image Processing: Principles and Applications*. USA: John Wiley and Sons. pp. 1–426. doi: https://doi.org/10.1002/0471745790

Ahmad, F. H. (2015). Efficient Facial Image Feature Extraction Method for Ethnicity Identification, M.Sc. Thesis. College of Commerce, University of Sulaimani, Sulaimani. pp. 1–71

Ahmed, A., Salam, B., Mohammad, M., Akgul, A., & Khoshnaw, S. H. A. (2020). Analysis coronavirus disease (COVID-19) model using numerical approaches and logistic model. *AIMS Bioeng.*, *7*(3), 130–146.

Cevallos, C., Ponce, H., Moya-Albor, E., & Brieva, J. (2020, July 1). Vision-Based Analysis on Leaves of Tomato Crops for Classifying Nutrient Deficiency using Convolutional Neural Networks. Proceedings of the International Joint Conference on Neural Networks, pp. 1-7. doi: https://doi.org/10.1109/IJCNN48605.2020.9207615

Chouvatut, V., Yotsombat, C., Sriwichai, R., & Jindaluang, W. (2015). Multi-view hand detection applying viola-jones framework using SAMME AdaBoost. Proceedings of the 2015-7th International Conference on Knowledge and Smart Technology, KST 2015. pp. 30-35. doi: https://doi.org/10.1109/KST.2015.7051476

Da'San, M., Alqudah, A., & Debeir, O. (2015). Face detection using Viola and Jones method and neural networks. 2015 International Conference on Information and Communication Technology Research, ICTRC 2015. pp. 40-43, doi : https://doi.org/10.1109/ICTRC.2015.7156416

Dey, S., Howlader, A., & Deb, C. (2021). MobileNet Mask: A Multi-phase Face Mask Detection Model to Prevent Person-To-Person *Transmission of SARS-CoV-2*, pp. 603–613. doi: https://doi.org/10.1007/978-981-33-4673-4_49

Hazim, N., Sameer, S., Esam, W., & Abdul, M. (2016). Face Detection and Recognition Using Viola-Jones with PCA-LDA and Square Euclidean Distance. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *7*(5) 371-377. doi: https://doi.org/10.14569/ijacsa.2016.070550

Hendra, T., Spolaor, R., & Chen, Z. (2019). A Compound Technique for Multiple Objects Detection Based on Markov Clustering Networks and Viola-Jones Algorithm. 2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP), pp. 459–463. doi: https://doi.org/10.1109/ICICSP48821.2019.8958601

Kolsch, M., & Turk, M. (2004). Robust hand detection. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings., FGR Vol. 4, pp. 614-619. doi: https://doi.org/10.1109/AFGR.2004.1301601

Kovalenko, M., Antoshchuk, S., & Sieck, J. (2014). Real-time hand tracking and gesture recognition using semantic-probabilistic network. Proceedings - UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014, pp. 269-274. doi: https://doi.org/10.1109/UKSim.2014.49

Mao, G. Z., Wu, Y. L., Hor, M. K., & Tang, C. Y. (2009). Real-time hand detection and tracking against complex background. IIH-MSP 2009 - 2009 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 905–908. doi: https://doi.org/10.1109/IIH-MSP.2009.133

Mohammad, M., Hicks, Y., & Kaloskampis, I. (2016). Video-based Road Detection Using Evolving GMMs and Region Enhancement. 11th International IMA Conference on Mathematics in Signal Processing, Birmingham, Dec 2016.

Muhammad, M. A. (2016). Video-based Situation Assessment for Road Safety. Ph.D. Thesis, Cardiff University, Cardiff, 1-187.

Murthy, C. B., Hashmi, M. F., Bokde, N. D., & Geem, Z. W. (2020). Investigations of object detection in images/videos using various deep learning techniques and embedded platforms-A comprehensive review. *Applied Sciences (Switzerland),* 10(9). doi: https://doi.org/10.3390/app10093280

Nguyen, V.-T., Le, T., Tran, T.-H., Mullot, R., & Courboulay, V. (2012). A method for hand detection based on Internal

Haar-like features and Cascaded AdaBoost Classifier. Conference: Proceedings of The Fourth International Conference on Communications and Electronics (ICCE 2012), pp. 608–613.
Verschae, R., & Ruiz-del-Solar, J. (2015). Object detection: Current and future directions. *Frontiers Robotics AI*, *2*(NOV). doi: https://doi.org/10.3389/frobt.2015.00029
Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1. CVPR 2001, pp. 1-9. doi: https://doi.org/10.1109/cvpr.2001.990517
Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., Yi, P., Jiang, K., Wang, N., Pei, Y., Chen, H., Yu, M., Huang, Z., & Liang, J. (2020). *Masked Face Recognition Dataset and Application*. arXiv preprint arXiv:2003.09093.
Yun, L., & Peng, Z. (2009). An automatic hand gesture recognition system based on Viola-Jones method and SVMs. 2nd International Workshop on Computer Science and Engineering, WCSE 2009, 2, pp. 72–76. doi: https://doi.org/10.1109/WCSE.2009.769