Video Analysis on Handwashing Movement for the Completeness Evaluation

Esa Prakasa
Computer Vision Research Group
Research Center for Informatics,
Indonesian Institute of Sciences (LIPI)
Bandung, Indonesia
esa.prakasa@lipi.go.id

Bambang Sugiarto
Computer Vision Research Group
Research Center for Informatics,
Indonesian Institute of Sciences (LIPI)
Bandung, Indonesia

***Abstract*—The pandemic situation of COVID-19 is still ongoing in every part of the world countries. Up to now, the medicine and the vaccination for curing the COVID-19 are not yet available. Since the vaccine and medication of COVID-19 are not yet available, the government in many countries are strongly socialising three kinds of new practices. The new habits are physical distancing, wearing a mask, and handwashing frequently. This paper focuses on assessing the completeness of handwashing. The handwashing frames are extracted from a video clip. The frames are then categorised into six-movement steps by using a deep learning-based algorithm. The frame of hand object is separated from the video by applying skin colour differentiation. Two kinds of experiments based on two different video sources are performed in the paper. Scoring simulation is also conducted in the research. The results show that the proposed method can give an excellent performance with accuracy values not less than** 90%.

Keywords: video analysis, handwashing, classification

I. Introduction

The coronavirus COVID-19 is still affecting all countries in the world. The COVID-19 pandemic was initially found in Wuhan China at the end of 2019. There is no sign the pandemic will end in near term even though there is good progress in vaccine development and decreasing cases in several countries. The new status found that the COVID-19 has infected 59,684,722 of the world population. A total of 41,289,731 people have been recovered, but 1,405,722 have died [1]. In Indonesia, COVID-19 has affected a total of 506,302 people, and 425,313 are successfully recovered. However, around 3.2% of the confirmed cases, 16,111 have died [2]. The Indonesian government is still struggling to fight COVID-19 pandemic, either as health and economic problems.

The vaccine and medicine development are currently still in progress. The development has been predicted will not complete in the current year. Therefore, preventive actions for minimising COVID-19 spread should be conducted by all parties. The government from many countries have strongly encouraged new practices that need to follow. There are three primary practices for the people during the COVID-19 pandemic. Those practices are to keep physical distancing, always wearing a mask in the public area, and handwashing frequently using either soap or hand sanitiser. The paper focuses on handwashing practise. The medical workers already perform the handwashing with the complete step since along ago. They need to do this procedure to ensure the purify of medical services. Their action has been proven to provide a healthy environment for the patient and the medical staff. However, the standardised handwashing step is not widely known by the ordinary people. The handwashing method needs to be more disseminated and socialised. Therefore, the paper proposes a technique based on video scene analysis to assess the handwashing completeness. The availability of this system could give a lot of benefits informing the handwash quality. People can get feedback directly from the system on how they did handwashing. If the score is high, they can be considered as people with adequate information on suggested handwashing steps. In the other hand, when the handwashing score is low, then the people can discern their understanding is not yet correct. Therefore, they can try to improve their handwashing capability. In the future, the proposed scoring system might be integrated into the access gate system with certain cleanse level, despite functioned as a dissemination tool as in the current presentation.

As depicted in Figure 1, there are six-movement steps of a complete handwashing. This step is encouraged by the World Health Organization as a series of step for ensuring the completeness of handwashing. The stage is initialised by rubbing hands palm to palm. The hand is moved in circle direction to rub the palm surfaces. After the first step has been completed, the next step is rubbing left dorsum with the right palm and vice versa. The interlaced fingers are also rubbed during this stage. The third stage is cleaning palm to palm surfaces. The interlaced fingers are also needed to be cleaned in the current step. The fourth step is conducted by opposing the backs of the interlocked fingers to the palm surface of the other hand. Next, the fifth stage is performed by the rotational rubbing of the left thumb clasped in right palm and vice versa. The last movement is completed by rotational rubbing, backwards and forwards with the clasped finger of the right hand and in the left palm and vice versa. All of these movements need to complete in around 20 seconds

The paper consists of five sections, and the first section describes the research background, problems, and objective

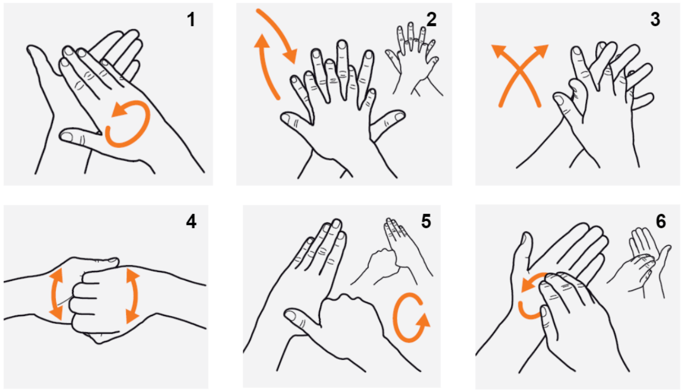


Fig. 1. Six basic movements of handwashing [3].

of the proposed method. Previous works are discussed in the second section. The methodology, including data collection, algorithm design, and classification techniques, are given in the third section. The fourth section elaborates the implementation and its results. The algorithm performance is also presented. Finally, the conclusion and acknowledgement end the paper.

II. Previous Works

Several works have been carried out to analyse the perfect handwashing. Wang *et al.* [4] have studied the sensor armband accuracy to meet the WHO hand rub and handwash standards. The research participants are instructed to wear the Myo armband on their either forearm or arm. The movement patterns created in performing handwashing are recorded and classified. The Xtreme Gradient Boosting (XGBoost) and Energy Divisive algorithms are used to analyse the handwashing completeness. Two cross-validation methods, leave-one-session-out (LOSO) and leave-one-participant-out (LOPO) are used to measure the model accuracies. The method objectives are to create models of user-dependent and user-independent, respectively. According to the evaluation results, the hygiene quality level can be accurately determined based on the handwashing and hand rubbing poses. A single armband can be worn in the upper arm area to obtain acceptable performance. However, this result could be improved by moving the armbands location from the upper arm to the forearm. Improvement can also be realised by using double armbands.

Llorca *et al.* [5] developed a new method based on computer vision techniques for measuring handwashing steps. The handwashing guidance is considered as the reference steps. The hand regions are segmented by using a combination of colour and motion analysis. The single multi-modal particle filter (PF) and the k-means clustering are collaborated to track hands and arms motions. Meanwhile, the hand gestures are categorised by applying a support vector machine classifier. Hanafi *et al.* [6] developed a simple mobile application engaging augmented reality, namely Mobile Augmented Reality Hand Wash (MARHw). This application will educate the urban and rural community on a notable effect of basic infection prevention practices peculiarly on handwashing guidance to prevent COVID-19 infection. This paper showed that the participant is required much more time to likely familiar with the steps in MARHw interfaces using statistical significant.

Ameling *et al.* [7] have also implemented computer vision algorithms to obtain an automated system of the handwashing measurement. The system is named as SureWash and can ideally classify handwashing poses. The system provides real-time feedback to the user on the quality of the accomplished handwashing. The system performance is evaluated by comparing agreement between two human observers and the system. The evaluation results show the high agreement among the raters that proves the accuracy and validity of the proposed method. All of these previous works can give a global view of the potential solution for assessing the handwashing completeness.

III. Methodology

A. Data Collection

Handwashing frames were extracted from a video provided by the Ministry of Health, Republic of Indonesia. The video demonstrates practising handwashing in six steps. The recorded video is stored in RGB colour format and available in the internet[8][9]. This clip will be defined as MOHRI video. A total of $1,647$ frames can be extracted from this single video. Figure 2 displays the six main steps of handwashing. The images are collected from extracted frames at index of 0360, 0510, 0720, 0850, 1160, and 1320.

The classification model will be developed by applying the convolutional neural network (CNN) architecture. The model need several dataset in the model development. Therefore, the extracted frames are divided into three groups for preparing the development stages. The groups are defined as training, validation, and testing dataset. The proportion of these groups are 70%, 20%, dan 10%, respectively. In each movement group, a total of 145 frames are collected. These frames are then randomly distributed into the three development groups. Random selection is obtained by applying permutation to the file indices. By following, the defined proportion, the number of each group are 120, 29, and 14. Therefore, the total frames used in the model development is $145 \times 6 = 870$.

B. Segmentation of Hand Region

Colour conversion from red, green, and blue (RGB) channels to hue, saturation, and value (HSV) channels is performed to obtain image component with high contrast on the skin human region. The best colour component in segmenting hand region is the hue channel. Image masking based on intensity thresholds is applied to create a binary image of hand region. Blur filter 5×5 is used to remove unwanted noises. The hand region is segmented by using the hue channel. The region will have the lowest intensity compared to the other regions. An intensity threshold is applied to the hue image to obtain the segmented hand region. Equation 1 is used to segment the hue channel image. The binary result is then inverted by subtracting the maximum intensity, 255, with the value of

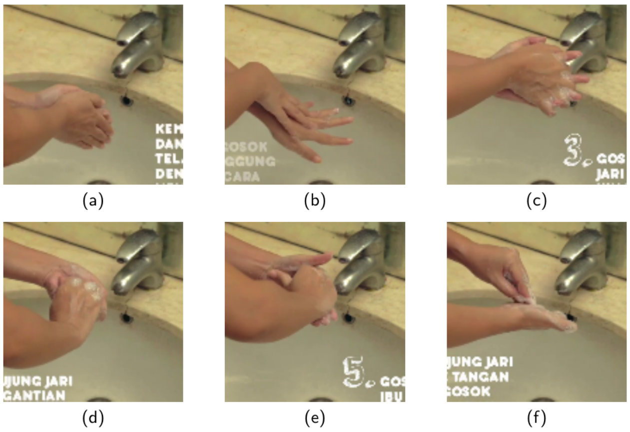


Fig. 2. Six main steps of handwashing. The step is started from figure (a) and ended at (f).

Fig. 3. The steps of hand region segmentation: (a) Original RGB image, (b) the extracted hue channel, and (c) segmented hand region.

the segmented image. This inversion process is formulated by Equation 2.

$$I_B\left(x,y\right) = \begin{cases} 255, I_H\left(x,y\right) \geq H_T \\ 0, I_H\left(x,y\right) < H_T \end{cases} \tag{1}$$

$$I_s\left(x,y\right) = 255 - I_B(x,y) \tag{2}$$

Variable I_H denotes the hue channel image extracted from the RGB colour image. H_T is written to symbolise the intensity threshold. Meanwhile, I_B denotes a binary image of the segmentation result. Equation 2 is required to flip the hand region becomes an area with high intensity and low intensity for the other part.

The H_T value is determined by averaging the intensities of the four sample areas at the image corners. Figure 4 displays four rectangle area at the image corners. The sample areas are named as C_1, C_2, C_3, and C_4. The threshold intensity, H_T, is computed by applying Equation 4. The dimension of input image is 108×108 pixels whereas the size of sample area is 10×10 pixels. The dimension size uses variable D to represent width and height of the sample area. The average intensity at k-th corner is calculated by using Equation 4. Variable $\overline{C}_k$ is used to denote the average intensity at the k-th corner.

$$H_T = 0.5 \times \left(\frac{\overline{C_1} + \overline{C_2} + \overline{C_3} + \overline{C_4}}{4}\right) \tag{3}$$

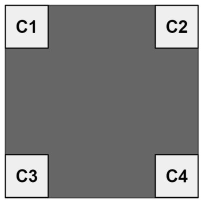


Fig. 4. Four corner samples for determining intensity threshold.

1	InputLayer, image size: 108x108
2	Conv2D, 64 filters, size: 3x3
3	Conv2D, 32 filters, size: 3x3
4	MaxPooling2D, size: 2x2
5	Conv2D, 32 filters, size: 3x3
6	Conv2D, 16 filters, size: 3x3
7	MaxPooling2D, size: 2x2
8	Dense, ReLU, 2,000 neurons
9	Dense, ReLU, 256 neurons
10	Dense, ReLU, 128, neurons
11	Dense, SoftMax, 6 neurons

Fig. 5. The CNN architecture for classifying hand movements.

$$\overline{C}_k = \frac{\sum_{i=0}^{D}\sum_{j=0}^{D} I_H\left[i_{0,k}+i, j_{0,k}+j\right]}{m \times n} \tag{4}$$

The reference index of each corner group, $i_{0,k}$ and $j_{0,k}$, depends on the corner locations. The index values can be described as shown in Equation 5 to 8. Variables M and N represent height and width of the input image, respectively.

$$k = 1; i_{0,1} = 0; j_{0,1} = 0 \tag{5}$$

$$k = 2; i_{0,1} = 0; j_{0,1} = N - D \tag{6}$$

$$k = 3; i_{0,1} = M - D; j_{0,1} = 0 \tag{7}$$

$$k = 4; i_{0,1} = M - D; j_{0,1} = N - D \tag{8}$$

C. *Architecture of Convolutional Neural Network*

Several neural networks layers are combined to create an architecture based on convolutional neural networks. This architecture will be used to train the prepared dataset to create a classifier model, as depicted in Figure 5. Two clusters of convolutional layers compose the architecture. Each group is formed by double two-dimensional convolutional layers and a single of a max-pooling layer. By applying this architecture, the total number of network parameters used is 19,014,054. The CNN model is set to compile using Adam optimiser. Since, the output consists of more than two options, the loss type is set to categorical cross-entropy.

Fig. 6. Data collection on the controlled background.

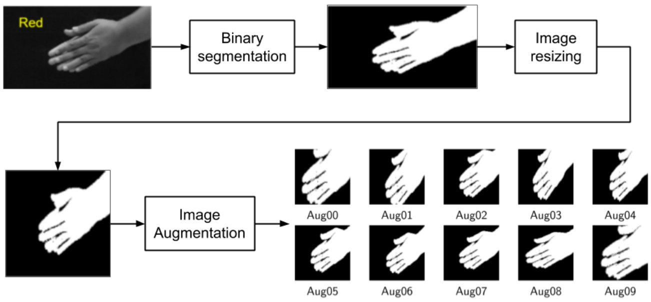


Fig. 7. Preparation of training dataset.

D. Data Preparation of Hand Model

The second experiment is conducted by collecting handwashing video in the controlled environment. Original images are collected from hand movement in front of a green background. This setting is selected to ensure the hand can be perfectly segmented from the background. Figure 6 depicts seven classes of the collected data. In this data collection, the seven classes represent six basic handwashing movement and another class of non-handwashing activity. The hand region is then segmented by using the intensity of the red channel of RGB colour space. The segmentation process provides binary images of hand poses. As shown in Figure 7, the binary image is then resized to obtain a square shape image. Augmentation algorithm is applied to the image dataset to increase data variation. The augmentation applies some geometric transformations at several parameter values—the transformation, including rotation, scaling, and translation. The augmentation process contributes to complete the sample number of each handwashing movement. Therefore, the image numbers will be balanced in each class at a total of 500 images.

IV. Result and Discussion

A. Implementation of the CNN Algorithm

The training stages is conducted by applying the frame dataset. The binary images as shown in Figure 3(c) are used as input images. Several parameters settings have been defined such as batch size, number of epoch, and the class mode. The batch size and number of epoch are set at 15 and 50, respectively. The class mode selects categorical option as the output type of the developed model. The architecture is trained using TensorFlow 1.10.0 [10] in Anaconda 4.8.3 environment [11]. The codes are written by using Python 3.6.5 [12] and additional libraries, Numpy 1.18.1 [13], [14] and OpenCV 3.4.1 [15]. The software packages are run in a computer with processor Intel® Core™ i7-8550U CPU @ 1.80 GHz, RAM 16.0 GB, and Microsoft Windows 10 operating system.

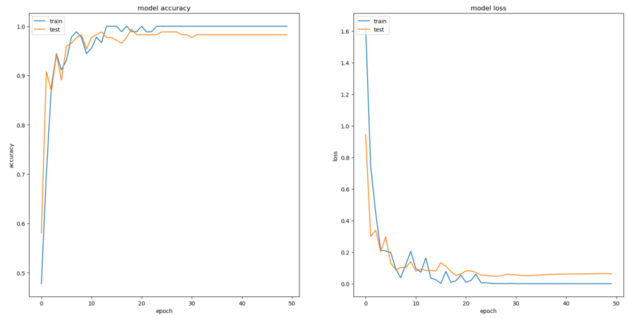


Fig. 8. The plot of model accuracy and loss obtained from the first experiment.

The training stage can be completed in 50 iteration steps. A good accuracy, 99.43%, is successfully achieved at the final step. By applying this model, the accuracy values are going to be higher than 90% after the fifth iteration. The accuracies have then fluctuated until reach the best accuracy at the endpoint of iteration. Once, the model wholly trained, the testing images are applied to observe the performance characteristics. The testing stage gives a confusion matrix and a table of performance values, as shown in Figure 9 and Table I, respectively. According to the test results, it can be found the classifier model can correctly differentiate the type of handwashing movements. Almost all of the movement class can be accurately detected. The experiment proved the proposed method could give high reliability. The final accuracy of the model is found at 98.81%. The average values computed from each movement class, including precision, recall, and $F1$-score are also higher than 98%. By referring the Cohen-Kappa score, the result is able to be categorised as almost perfect agreement [16].

The second experiment performs training in 200 iteration steps. The best accuracy of the training stages can be achieved at 98.89%. The accuracy goes to higher than 90%, starting from the 50th iteration. The testing images are then applied to the trained model. The testing results is described as a confusion matrix in Figure 11 and listed performance parameters in Table I. The parameter values of the second experiment are lower compared to the first experiment. This results due to additional augmentation to the training dataset. Therefore, the shape and position transformation are more provided compared to the previous experiment. According to the Cohen-Kappa score, the classification result is still considered as almost perfect agreement.

B. Classification on Handwashing Video

The trained model obtained from the second experiment is then tested by using the handwashing video. The image preprocessing and classification using CNN is performed

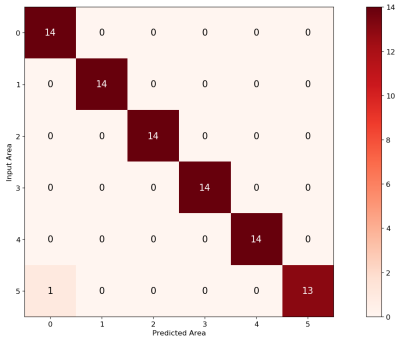


Fig. 9. The confusion matrix of the testing dataset in the first experiment.

TABLE I
THE VALUES OF PERFORMANCE PARAMETERS

No	Parameter	Experiment 1 (%)	Experiment 2 (%)
1	Accuracy	98.81	95.14
2	Average precision	98.89	95.16
3	Aveage recall	98.81	95.14
4	Average $F1$-score	98.81	95.15
5	Cohen-Kappa score	98.57	94.33

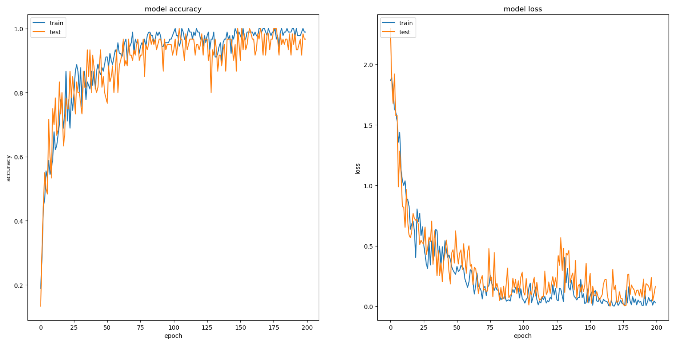


Fig. 10. The plot of model accuracy and loss obtained from the second experiment.

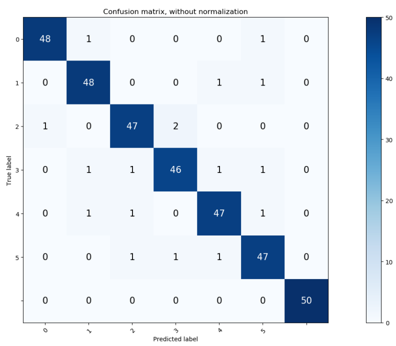


Fig. 11. The confusion matrix of the testing dataset in the second experiment.

frame by frame. The decision regarding the hand movement class is counted as the number of the frame. The accumulated number of classified frames are used as an indicator of the completeness of the handwashing action. The time duration in each pose is calculated by using an accumulation number

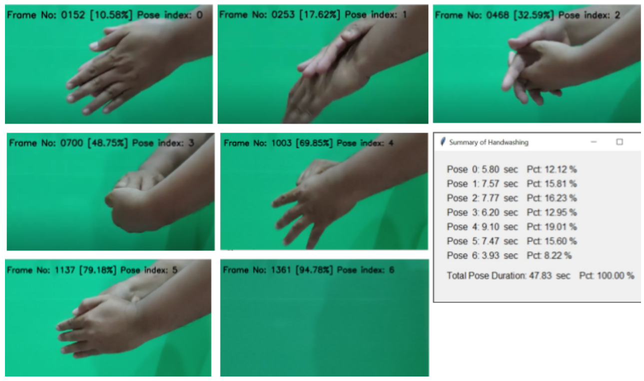


Fig. 12. The confusion matrix of the testing dataset in the second experiment.

of the classified frame. Figure 12 illustrates the frames that correctly detected according to their pose classes. The frame without hand object can also be identified as the 7th class. Total duration of the handwashing video is 47.83 s, whereas around 92% (43 s) is spent by handwashing movements.

C. Prototype of Scoring System

An algorithm that simulates the scoring process is also created. The scoring process can be conducted if the movement of handwashing is already correctly classified. The simulation is conducted by developing an algorithm that can capture a video in real-time and directly compute the duration of the movement steps. To represent handwashing movements, six object with different colours are used as the observed objects. The colours are yellow, black, green, blue, violet, and red. An additional class is then added, to represent a condition where the object is not available. This condition is named as "No Action". Therefore, seven classes are considered in the scoring simulation.

The object images are acquired by using the laptop webcam. The camera can record colour video, therefore RGB (Red, Green, Blue) channels are processed to obtain representative features. The RGB colour features are collected from the centre area of the object surfaces. The K-means clustering is applied to find the colour centroid in term of red, green, and blue values. Examples of the class frames are displayed in the Figure 13. A clip video showing the objects with different colour is categorised by the colour centroid of the K-means clustering. The Euclidean distance is used to decide the class of the observed frame. Several series of the frames are displayed in Figure 14. The video shows the real-time classification for scoring simulation can be found in the internet [17].

Figure 15 displays summary of time duration in the scoring simulation. The action types are represented by colour names. The duration time is counted by using the number of frames. It can be defined that there is around 15 second for collecting a single image. The step duration is computed by applying multiplication between number of frames and duration each frame.

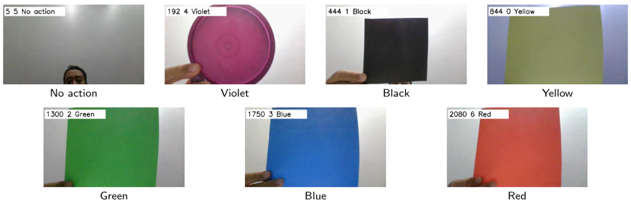


Fig. 13. Seven classes of the scoring simulation.

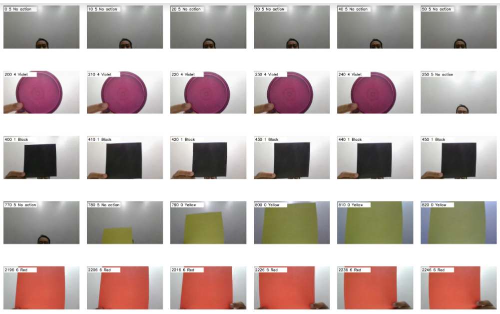
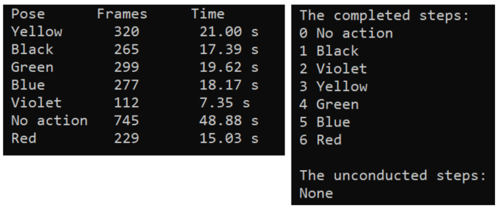


Fig. 14. Examples of frames extracted from the scoring simulation video.



Pose	Frames	Time
Yellow	320	21.00 s
Black	265	17.39 s
Green	299	19.62 s
Blue	277	18.17 s
Violet	112	7.35 s
No action	745	48.88 s
Red	229	15.03 s

The completed steps:
0 No action
1 Black
2 Violet
3 Yellow
4 Green
5 Blue
6 Red

The unconducted steps:
None

Fig. 15. Summary of time duration in the scoring simulation

V. Conclusion

The initial study has proposed a method for classifying the handwashing movement. The algorithm uses handwashing video and a classifier based on convolutional neural network architecture. Handwashing steps are extracted in the form of video frames and then individually classified. Image pre-processing algorithm has been applied to prepare the input data for the classification algorithm. The training and testing stages of two experiments can give strong accuracy values. The first experiment provides accuracies of training and testing stages at 99.43% and 98.81%, respectively. Meanwhile, the training and testing accuracies of the second experiment are 98.81% and 95.14%. Scoring simulation is also studied in the paper by observing the object with different colours. The time duration of the object appearance can be automatically determined. The simulation outcome shows that the handwashing classification and the scoring method are potentially combined to obtain a complete handwashing scoring system.

Acknowledgement

The authors would like to thank the Deputy for Engineering Sciences, Indonesian Institute of Sciences (LIPI) for funding the research and the paper publication. The study is part of the LIPI contribution to overcoming COVID-19.

Author Contribution

Esa Prakasa is the main contributor of the paper. Esa Prakasa prepared the experiment, implementing the algorithm, analysed the result and writing the manuscript. Bambang Sugiarto is the contributor member of the paper. Bambang Sugiarto reviewed and reported the previous works related to the paper topic.

References

[1] Worldmeters, "Covid-19 coronavirus pandemic," 4 August 2020. [Online]. Available: https://www.worldometers.info/coronavirus/

[2] G. T. P. P. Covid-19, "Peta sebaran covid-19," 4 August 2020. [Online]. Available: https://covid19.go.id/peta-sebaran

[3] WHO, "The members of the infection control programme-hand hygiene: Why, how, and when," 2009. [Online]. Available: https://www.who.int/gpsc/5may/Hand_Hygiene_Why_How_and_When_Brochure.pdf

[4] C. Wang, Z. Sarsenbayeva, X. Chen, T. Dingler, J. Goncalves, and V. Kostakos, "Accurate measurement of handwash quality using sensor armbands: Instrument validation study," *JMIR mHealth and uHealth*, vol. 8, no. 3, p. e17001, 2020.

[5] D. F. Llorca, I. Parra, M. n. Sotelo, and G. Lacey, "A vision-based system for automatic handwashing quality assessment," *Machine Vision and Applications*, vol. 22, no. 1, p. 219–234, 2011.

[6] H. F. Hanafi, M. H. Wahab, K.-T. Wong, A. Z. Selamat, M. H. M. Adnan, and F. H. Naning, "Mobile application to guide community to ameliorate hand washing effectiveness to oppose covid-19 disease," *International Journal of Integrated Engineering*, vol. 12, no. 5, p. 217–223, 2020.

[7] S. Ameling, J. Li, J. Zhou, A. Ghosh, G. Lacey, E. Creamer, and H. Humphreys, "A vision-based system for hand washing quality assessment with real-time feedback," in *BioMed 2010*, 2010.

[8] M. of Health Republic of Indonesia, "6 langkah cuci tangan pakai sabun," 2020. [Online]. Available: http://promkes.kemkes.go.id/video-ilm-6-langkah-cuci-tangan-pakai-sabun-ctps

[9] ——, "6 langkah cuci tangan pakai sabun," 2020. [Online]. Available: https://youtu.be/V4gUbrqMWk4

[10] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[11] Anaconda Software Distribution, "Anaconda." [Online]. Available: https://anaconda.com

[12] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

[13] T. E. Oliphant, *A guide to NumPy*. Trelgol Publishing USA, 2006, vol. 1.

[14] S. Van Der Walt, S. C. Colbert, and G. Varoquaux, "The numpy array: a structure for efficient numerical computation," *Computing in Science & Engineering*, vol. 13, no. 2, p. 22, 2011.

[15] OpenCV, "Open source computer vision library," 2015. [Online]. Available: https://opencv.org/

[16] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica: Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[17] E. Prakasa, "Colour classification based on rgb values," 2020. [Online]. Available: https://youtu.be/UpOz6CLpIpI