Journal of Computer Science and Cybernetics, V.37, N.3 (2021), 279–289
DOI 10.15625/1813-9663/37/3/15962

REAL-TIME FACE MASK DETECTION WITH FACIAL TEMPERATURE MEASUREMENT FOR COVID-19 INDOOR MONITORING SYSTEM

ARI AHARARI[1,*], JAIR MINORO ABE[2], KAZUMI NAKAMATSU[3]

[1] *Department of Computer and Information Sciences, SOJO University, Japan*

[2] *Graduate Program in Production Engineering, Paulista University, Brazil*

[3] *School of Human Science and Environment, University of Hyogo, Japan*

Abstract. The coronavirus disease (COVID-19) was first detected in 2019. To ensure protection from the disease and to prevent its transmission, wearing a face mask in public areas is strongly recommended. This work presents an indoor monitoring system that supports real-time face mask detection and facial temperature measurement. The proposed system employs ultrasonic sensors to detect people, and uses a CNN network (i.e., MobileNetv2) to detect face masks. Additionally, the proposed system uses Grid-Eye sensors to estimate the facial temperature of people. The proposed method was evaluated using a personal dataset and it achieved a detection accuracy of 98.8%, outperforming existing baseline models. In addition, the inference time was approximately 16 FPS on a CPU machine.

Keywords. Deep learning; Face mask detection; COVID-19; Convolutional neural network (CNN); Facial temperature measure.

1. INTRODUCTION

The coronavirus disease (COVID-19) was first detected in 2019 and it was declared as a global pandemic by the World Health Organization (WHO) in 2020 [1]. The global pandemic has severely affected the public health worldwide. The disease transmits through close contact, particularly in crowded environments. Therefore, several countries have enforced rules such as social distancing and wearing face masks in public areas. The WHO and the Centers for Disease Control and Prevention (CDC) have suggested that people older than two years of age should wear a face mask in public areas to prevent the transmission of COVID-19. Prior to the COVID-19 pandemic, people wore masks to protect their health from seasonal diseases such as influenza or from air pollution. Wearing a face mask reduces the transmission and spreading rate, especially when other social distancing measures [2] are difficult to maintain.

Dedicated to Professor Phan Dinh Dieu on the occasion of his 85th birth anniversary.

*Corresponding author.
E-mail addresses: aharari@cis.sojo-u.ac.jp (A. Aharari); jairabe@uol.com.br (J.M. Abe); nakamatu@pf7.so-net.ne.jp (K. Nakamatsu).

It was reported that more than 116 million people were infected by COVID-19 across 188 countries [3]. People and governments are facing extraordinary challenges and risks due to the coronavirus pandemic in several countries. In many countries, people are forced by laws to wear face masks in public. These laws and rules were developed to grow in cases and deaths in many countries exponentially. Surveillance has been performed by authorities to ensure that people are wearing face masks in crowded public areas, buildings, restaurants, shopping malls. However, monitoring and measuring the body temperature of people in public areas has become increasingly complex. An approach can be applied to overcome this difficulty by integrating body temperature detection and an artificial intelligence face mask detection system. The objective of this study is to detect face masks in an image and measure the skin temperature. In this paper, we propose a face mask detection model based on computer vision and deep learning. The proposed model can be integrated with a USB camera to detect whether or not a person is wearing a face mask to prevent the transmission of COVID-19. The model integrates deep learning and classical machine learning techniques with OpenCV, Tensorflow, and Keras. Additionally, deep transfer learning was applied for feature extraction and it was combined with a classical machine learning algorithm. The models were compared to determine a suitable algorithm that achieves a higher accuracy and requires lesser time for training and detection.

2. RELATED WORKS

A crucial aspect of computer vision is object detection in an image. Most related work focuses on face detection applying computer vision and recognizing the people who are not wearing face masks to decrease the transmission and spreading of the COVID-19. It has been proven that wearing a face mask reduces the transmission of COVID-19. This section reviews recent research papers for applying representative works related to face mask detection. Face mask detection techniques focus on face construction and face recognition based on conventional machine learning techniques. The Viola-Jones face detector is an extensively used model which adopts boosted cascade with simple Haar features [4]. Li et al. [5] proposed a multi-view face detector that adopted the surf features during training and testing processes inspired by the Viola-Jones face detector. In [6], a face detector was proposed to efficiently detect faces with an ensemble of optimized decision trees. Faces can be detected quickly by comparing pixel intensities in the internal nodes. Liao et al. [7] proposed a face detector that utilized the scale-invariant and normalized pixel difference of image feature. A single soft-cascade classifier was adopted for efficient face detection. Furthermore, a few techniques have proposed explicitly modelling the structure or deformation of faces with deformable part model (DPM). For example, Zhu and Ramanan [8] proposed a tree-structured model for face detection, simultaneously estimating face poses and localizing facial landmarks. Mathias et al. [9] trained a DPM-based face detector with 26000 faces from annotated facial landmarks in the wild (AFLW), which achieved an average precision of 97.14% on AFW [8]. Chen et al. [10] presented a face detector by jointly learning detection and alignment in a unified framework by observing that aligned face shapes can provide better face classification features. In [11], a model was proposed to jointly handle face detection and keypoint localization using hierarchical DPM. Typically, DPM-based face detectors achieve impressive accuracy. However, they are susceptible to high computational costs due to DPM usage.

CNN-based face detectors directly learn face representations from data which is different from that of boosting- based and DPM-based approaches [12, 13]. It adopts a deep learning paradigm [14, 15, 16] to detect the presence of a face in a scanning window. For example, Li et al. [17] proposed cascade CNN, a boosted exemplar-based face detector. Farfade et al. [13] finetuned the AlexNet [18] to obtain a multi-view face detector trained on 200000 positive samples and 20 million negative samples. Yang et al. [19] proposed a face detector that utilized the feature aggregation framework [20] while the features were generated through CNN. In [21], the faceless of a window was assessed with an attribute-aware CNN, and occlusions were considered to generate face proposals. A strong capability of this method demonstrates pose variation and detecting faces with severe occlusion. Zhu et al. [22] proposed contextual multi-scale region-based convolution neural network (CMS-RCNN) for face detection under unconstrained conditions. Li et al. [23] presented a learning framework for face detection in the wild by integrating a 3D face model in an end-to-end multitask and CNNs. A grid loss layer for CNNs was proposed by Opitz et al. [24] to deal with partial occlusion during face detection. Additionally, it was capable of minimizing the error rate on sub-blocks of a convolution layer. Chen et al. [25] proposed a cascaded CNN called supervised transformer network to address the challenge of large pose variations in real-world face detection. Ranjan et al. presented Hyperface for simultaneous face detection, landmarks localization, pose estimation, and gender recognition using CNNs [26]. It appropriately utilized the synergy among the tasks, which increased its individual performances.

3. PROPOSED SYSTEM

A real-time framework was proposed to detect a person not wearing a face mask or with a facial temperature greater than 37.5°C. In the proposed indoor monitoring system, the detection person module is constantly at stand-by and if the target monitored person walks within a distance less than 60cm, it moves to the next module.

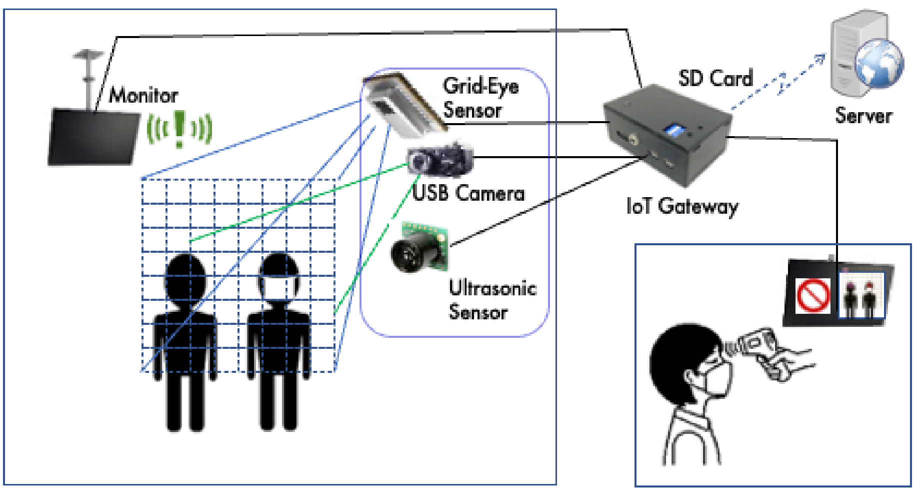

Figure 1: A framework of the proposed system

Skin temperature measurement is commonly used to explore the interaction between human thermophysiology and the external environment. While the core temperature is endothermic and is strictly regulated by the brain, the skin temperature is exothermic since it is affected by the environment and the "dual-thermic" thermoregulation ability. During heat stress, peripheral vasodilation increases blood flow in the skin with a consequent increase in temperature and heat dissipation. Conversely, peripheral vasoconstriction leads to a decrease in skin temperature and heat transfer to the environment during cold stress. Therefore, in the case of abrupt changes in environmental conditions, it is advisable to wait for an adequate stabilization time to attain a steady-state before the measurement.

A USB camera monitors the people and the captured images are fed into a face mask detection module to detect a person without a face mask in the image. Simultaneously, the facial temperature module measures the facial temperature of the people. If a person without a face mask or with a high facial temperature is detected, the information is transmitted to the monitor and an alert is displayed to ensure that the authorities perform necessary actions. The face-captured images are automatedly named by the date, time, and facial temperature and are saved in an SD card inside the IoT gateway. This information can be downloaded offline or can be automatically forwarded to the server. The framework of the proposed system is depicted in Figure 1.

A diagram of the proposed system with applied modules is shown in Figure 2. In the following sections, the detection of people, face mask detection, and temperature measurement are explained in detail.

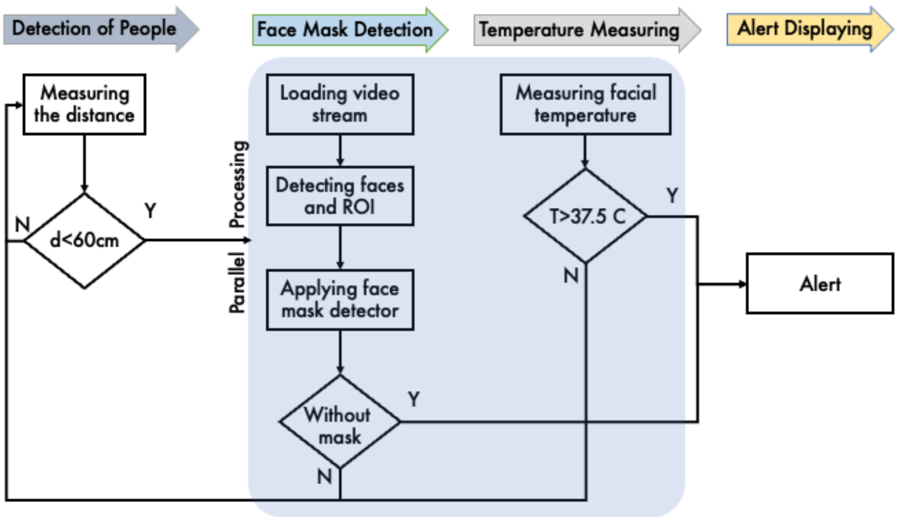

Figure 2: A diagram of the proposed system with applied modules

3.1. Detecting of people using ultrasonic sensors

In the proposed system, an ultrasonic sensor was used to detect the people in front of the system. The ultrasonic sensor can monitor a the movement of a person and the direction of the movement due to its high sensitivity and ability to detect people. Advertisements can be displayed on the monitor when the system is non-operational. The ultrasonic sensor is used to switch from the advertisement page to the measuring system while detecting people.

The detection target area was set to 60 cm in the proposed system. The detection area to the 2.54cm diameter dowl generally represents a reliable area to detect people with an ultrasonic sensor. Figure 3 shows the sample results for the measured beam pattern to detect people using LV-MaxSonar- EZ4 [27]. The detection pattern is shown in Figure 3 for dowels of varying diameters on a 30 cm grid placed in front of the sensor.

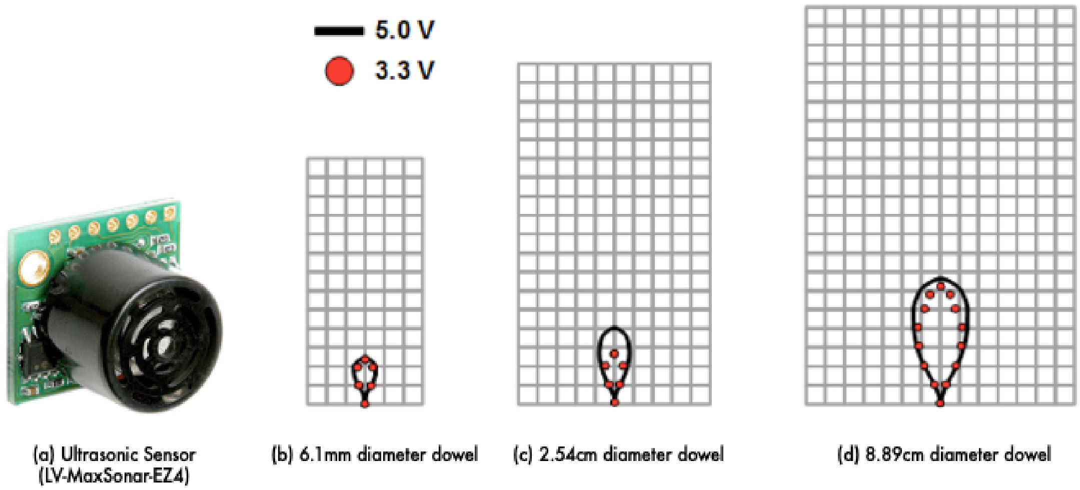

Figure 3: Sample results of the measured beam pattern using LV-MaxSonar-EZ4

3.2. Face mask detection

Deep learning techniques learn various important nonlinear features from the given samples. Therefore, the deep learning model enables the prediction of previously unseen samples. We collected 6450 face images with masks and 6150 face images without masks to train the face mask detection model. The images in the dataset were resized to 260×260 pixels during the preprocessing step. A sample of face dataset images is shown in Figure 4. The face dataset image was applied to the trained model and the model was serialized during the training step. Haar feature-based cascade [30] was applied to the video stream captured by the USB camera to detect faces in the images and extract the region of interest (ROI) of each face. Subsequently, the trained model was applied to the face ROI images to detect a face with a mask or without a mask. Finally, the results demonstrated that the green rectangular frame individually interprets the a face detected with a mask and the red rectangular frame to detect a face without a mask. Python programming was used to develop the detector model. Additionally, the model used Facemasknet architecture with eight layers for training.

The input layer accepts the training dataset images, and the rectified linear unit was used to develop the convolutional neural network (CNN). MobileNetV2 is a CNN architecture that performs well on mobile devices. It is based on an inverted residual structure where the residual connections are between the bottleneck layers. The intermediate expansion layer uses lightweight depthwise convolutions to filter features as a source of non-linearity. The architecture of MobileNetV2 [31] contains the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers. We used the trained MobileNetV2 model to detect whether or not a person is wearing a mask. The MobileNetV2 architecture is a highly efficient architecture applied to embedded devices with limited computational capacity such as Raspberry Pi. The training model compiled in this study was evaluated on the test dataset. The complete model was applied to real-time images to detect if a person was wearing a face mask or not. The primary goal of the proposed system is to detect a person with high facial temperature or without a face mask. The above-mentioned face mask detector identifies if an image consists of a person without a face mask. If such a person is detected, the information is transferred to the monitor to display an alert. Subsequently, this will enable an authorized person to perform necessary actions such as measuring the body temperature with high accuracy devices or deny passage of the person without a mask.

(a) Sample images with mask (b) Sample images without mask

Figure 4: A Sample of face dataset images

3.3. Facial temperature measure using grid-eye sensor

A Grid-Eye sensor was used in the proposed system, which is an infrared thermal sensing-based sensor with an image mapping device shown in Figure 5. This sensor enables contact-less temperature measurement in two-dimensional (8×8) 64 pixels areas. The sensor can be patched to other devices easily connected via I2C, enabling fast and efficient communication. It can detect human presence and body temperature without any contact with 64 thermopile elements. The lenses etched on top of the silicon wafer provide horizontal and vertical angles of view (60 degrees each). The Grid-Eye sensor has a compact size (11.6mm×8mm×4.3mm) which is suitable for application in portable devices. The Bluetooth module, MEMS sensor, lens, and I2C interface cover a floor of 2.5m×2.5m when installed 3m above the floor. The output data is displayed as a color thermal image data according to the 64 measured temperatures.

Grid-Eye sensor has a temperature measurement range of -20°C–100°C, and it has a measurement accuracy of ±3°C. Figure 6(b) shows a raw thermal image obtained from Grid-Eye sensors with a range of 50 cm. Bilinear interpolation was applied to remove the noise from the raw thermal image. It is a resampling method that uses the nearest pixel values' distance-weighted average to estimate a new pixel value. The applied bilinear interpolation thermal image is shown in Figure 6(c).

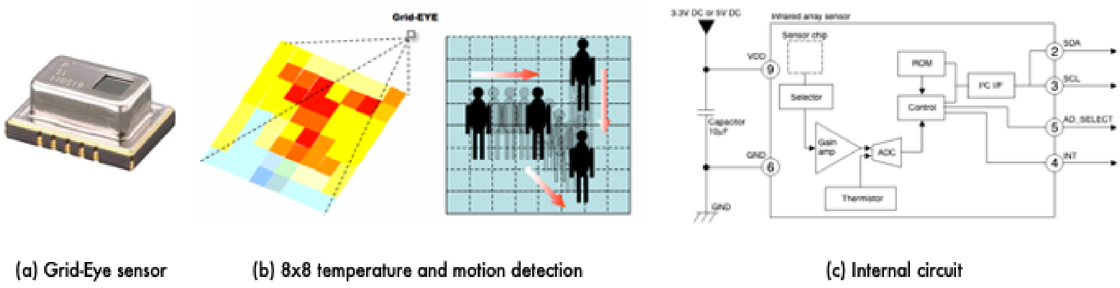

Figure 5: Outline of the Grid-EYE sensor (Panasonic)

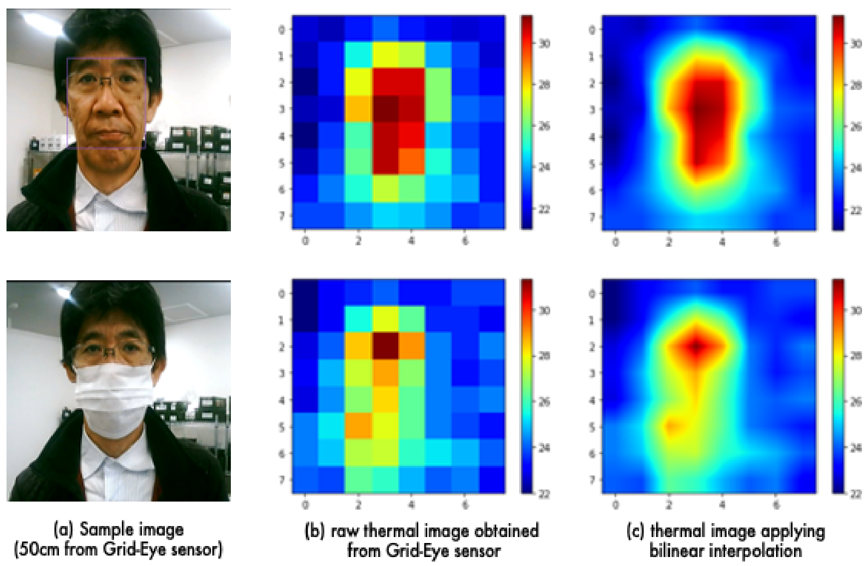

Figure 6: A Sample of a thermal image obtained using the Grid-Eye sensor

It can be observed from Figure 6 that the area near the eye is susceptible to noise and it is the hottest spot measured on the face in both face images with and without a mask. The temperature in this area was in the range of 34–36°C. Therefore, the average of infrared array areas with high temperature pixels can measure the facial temperature. The average normal body temperature is approximately 37°C. Recent studies have demonstrated that the average body temperature is in the range of 36.1–37.2°C. The body temperature of a healthy person fluctuates throughout the day, being cooler in the morning and warmer in the afternoon. The proposed system focuses on monitoring people with a facial temperature greater than 36°C through an alert on the display.

4. EXPERIMENTAL RESULTS

We collected 6450 face images with masks and 6150 face images without masks to train the face mask detection model. The images in the dataset were resized to 260×260 pixels in the preprocessing step. Python programming was used to develop the facemask detector model. To evaluate the accuracy of the trained model, the initial learning rate was set to

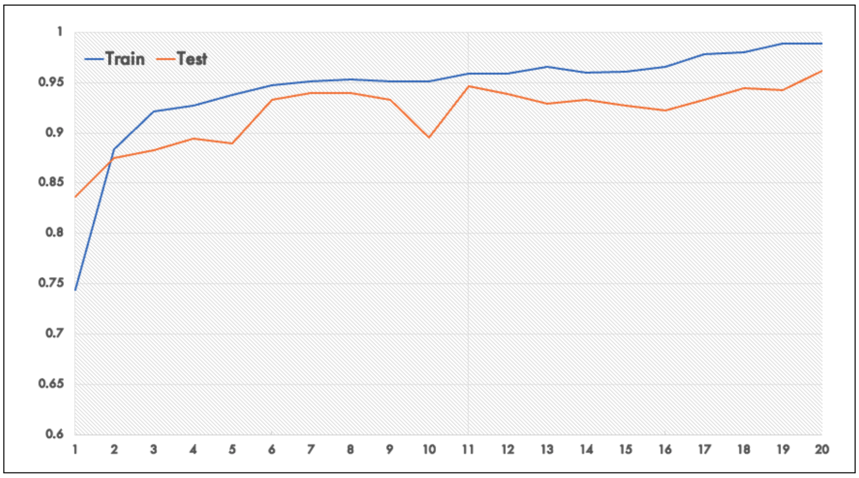

Figure 7: The accuracy and iteration curves of the proposed model

1e-4, and the number of training epochs was 20. The trained model was evaluated based on the test image dataset. The accuracy and iteration curves are plotted in Figure 7. The results demonstrate an accuracy of 98.8% on the test image dataset.

Furthermore, the accuracy of face mask detector model was compared with that of existing models shown in Table 1. The comparison results demonstrated that the proposed model achieved a better accuracy compared with that of other models. The average running time of the proposed method was 0.02 s for a single face when implemented with an Apple M1 and 7 core GPU with 8 GB memory.

Table 1: A comparison with other existing models

Method	Model	Accuracy	Running Time
Qin et al. [28]	SRCNet	98.7%	0.03
Khandelwal et al. [29]	MobileNetV2	97.6%	0.02
Proposed method	Facemasknet	98.8%	0.02

The proposed indoor monitoring system was tested by installing it at the entrance of a cafeteria at a university. The system was operated for three weeks, and during this period, the hardware and software reliability was confirmed. Figure 8 shows the proposed indoor monitoring system set up for evaluation.

5. CONCLUSIONS

The present paper proposed an online monitoring system to detect a person with high facial temperature or without a face mask. The proposed face mask detector identified if an image consisted of a person without a face mask. If such a person was detected, the information was transferred to the monitor to display an alert. Subsequently, an authorized person might perform necessary actions such as measuring the body temperature with high accuracy devices or deny passage to a person without a mask. The accuracy of the face mask detector model was 98.8%, and was higher than that of existing models. Additionally, the proposed indoor monitoring system to was evaluated for three weeks to verify the reliability of the system in real-time (16 FPS). The proposed monitoring system can be used in crowded

Figure 8: The proposed indoor monitoring system

places like bus stops, mall entrances, schools, and universities. Hence, appropriate measures performed by authorities through detection of people without face masks or with high facial temperature might reduce the transmission of COVID-19.

ACKNOWLEDGMENT

The authors would like to thank Daisuke Higashida, Masayoshi Inada, Takashi Hashimoto for their great support and advice during this work. This work was granted by Kumamoto City FY2020.

REFERENCES

[1] S. Feng, C. Shen, N. Xia, W. Song, M. Fan and B. J. Cowling, "Rational use of face masks in the COVID-19 pandemic," *Lancet Respir Med*, vol. 8, no. 5, pp. 434–436, 2020.

[2] C. Sun and Z. Zhai, "The efficacy of social distance and ventilation effectiveness in preventing COVID-19 transmission," *Sustainable Cities and Society*, vol. 62, Article 102390, 2020. https://doi.org/10.1016/j.scs.2020.102390

[3] WHO Coronavirus Disease (COVID-19) Dashboard. [Online]. Available: https://covid19.who.int/ (Accessed March 10, 2021).

[4] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 511–518, 2001. Doi: 10.1109/CVPR.2001.990517.

[5] J. Li and Y. Zhang, "Learning SURF cascade for fast and accurate object detection," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3468–3475, 2013.

[6] N. Markus, M. Frljak, I. S. Pandzic, J. Ahlberg, and R. Forchheimer, "A method for object detection based on pixel intensity comparisons organized in decision trees," *CoRR 2014*, 2014.

[7] S. Liao, A. Jain, and S. Z. Li, "A fast and accurate unconstrained face detector," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, vo. 2, pp. 211–223, 2016.

[8] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2879–2886, 2012.

[9] M. Mathias, R. Benenson, M. Pedersoli, and L. V. Gool, "Face detection without bells and whistles"," *European Conference on Computer Vision*, vol. 8692, pp. 720–735, 2014.

[10] D. Chen, S. Ren, Y. Wei, X. Cao, and J. Sun, "Joint cascade face detection and alignment," *European Conference on Computer Vision*, vol. 8694, pp. 109–122, 2014.

[11] G. Ghiasi and C. C. Fowlkes, "Occlusion coherence: Localizing occluded faces with a hierarchical deformable part model," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1899–1906, 2014.

[12] C. Zhang and Z. Zhang, "Improving multiview face detection with multi-task deep convolutional neural networks," *IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs*, pp. 1036–1041, 2014.

[13] S. S. Farfade, M. J. Saberian, and L. Li, "Multi-view face detection using deep convolutional neural networks," *ACM International Conference on Multimedia Retrieval*, pp. 643–650, 2015.

[14] R. Girshick, "Fast r-cnn," *IEEE International Conference on Computer Vision*, pp. 1440–1448, 2015.

[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, 2014.

[16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, pp. 91–99, 2015.

[17] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5325–5334, 2015.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Proc. of the 25th International Conference on Neural Information Processing Systems*, vol. 1, pp. 1097–1105, 2012.

[19] B. Yang, J. Yan, Z. Lei, and S. Z. Li, "Convolutional channel features," *IEEE International Conference on Computer Vision*, pp. 82–90, 2015.

[20] P. Dollár, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1532–1545, 2014.

[21] S. Yang, P. Luo, C. Loy, and X. Tang, "From facial parts responses to face detection: A deep learning approach," *IEEE International Conference on Computer Vision*, pp. 3676–3684, 2015.

[22] C. Zhu, Y. Zheng, K. Luu, and M. Savvides, "CMS-RCNN: contextual multi-scale region-based CNN for unconstrained face detection," *Book: Deep Learning for Biometrics*, 2017.

[23] Y. Li, B. Sun, T. Wu, and Y. Wang, "Face detection with end-to-end integration of a convnet and a 3D model," *European Conference on Computer Vison*, pp. 122–138, 2016.

[24] M. Opitz, G. Waltner, G. Poier, H. Possegger, and H. Bischo, "Grid loss: Detecting occluded faces," *European Conference on Computer Vison*, pp. 386–402, 2016.

[25] A. G. H. Dong Chen, F. Wen, and J. Sun, "Supervised transformer network for efficient face detection," *European Conference on Computer Vison*, pp. 122–138, 2016.

[26] R. Ranjan, V. M. Patel, and R. Chellappa, "Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 1, pp. 121–135, 2019.

[27] LV-MaxSonar-EZ4 datasheet. [Online]. Available: https://www.maxbotix.com/documents/LV-MaxSonar-EZ_Datasheet.pdf/

[28] B. Qin and D. Li, "Identifying facemask-wearing condition using image super-resolution with classification network to prevent covid-19," *Sensors*, vol. 20, no. 18, Article no. 5236, 2020.

[29] P. Khandelwal, A. Khandelwal and S. Agarwal, "Using computer vision to enhance safety of workforce in manufacturing in a post covid world," *arXiv preprint arXiv:2005.05287*, 2020.

[30] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, 2004.

[31] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.

Received on March 24, 2021
Accepted on September 05, 2021