2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Improve identity recognition with occlusion detection-based feature selection

Jaeyoon Jang, Hosub Yoon, Jaehong Kim

Abstract— **Image-based user recognition has become a technology that is now used in many ways. This is because it is possible to use only a camera without the need for any other device. In addition, due to the advantage of contactless, there are relatively many people who prefer it. However, normal recognition is not possible if some of the face information is lost due to the user's posture or the wearing of masks caused by the recent prevalent disease. This is because the information needed for recognition is largely lost due to obscurity. In some platforms, although performance is improved through incremental updates, it is still inconvenient. It takes a lot of time to update information about the occluded faces, and the overall accuracy is lowered. In this paper, we propose a method to respond more actively to these situations. First, determine whether the obscurity occurs, and improve the stability by calculating the feature vector using only a significant area when the obscurity occurs. By recycling the existing recognition model, without incurring little additional costs, it was confirmed the results of reducing the recognition performance drop in certain situations. Since only one layer has been added to the existing recognizer, there is no burden in terms of computational speed or memory. Using this technique, we confirmed the performance improvement of about 1.5% in the situation where some information is lost. Although performance is not dramatically improved, it is the biggest advantage that it can improve recognition performance by utilizing existing systems.**

I. INTRODUCTION

As image-based user authentication technology advances, face recognition technology is being used in various areas. The scope of use is also expanding to service markets such as contents that draw virtual data on the face and customized advertisements. In particular, the level of technology has increased to such a degree that it can be used in security areas such as financial settlement and identification cards. It is also used not only in terms of security but also in terms of entertainment. Despite the development of these technologies, there is still the problem remains. The biggest problem among them is that the entire face needs to be seen correctly for accurate certification. If you wear a mask due to cold weather or diseases such as the COVID-19, it is impossible to recognize normal because more than half of the face is covered. As well as accessories such as masks, it is impossible to recognize the normal identity even if part of the face area is cut off due to the user's posture in the interaction situation with the robot.

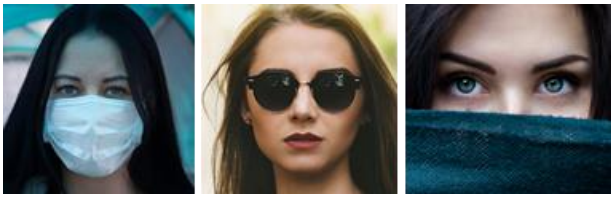

Figure 1. The case of occlusion on the face due to various factors.

For various reasons such as figure 1, facial information is lost, but we nevertheless want to gain strong recognition performance. Common studies so far have expected a single recognition model to be able to learn to be tough in a variety of environments and respond to all exceptional situations. Of course, due to the nature of deep learning, if only enough databases for various environments were built, we could expect that much performance, but realistically, it is impossible to gather data about all the situations. So we conducted a study on how to recognize identity more robustly in situations such as mask and screen truncation, using a previously learned recognition model. We first experimented to find areas that affect identity recognition by covering a certain percentage of the face area. In addition, we experimented to learn the recognizer to determine the situation of face screening. Then, if the screening occurs using only the unobscured area to extract the identity characteristics feature vector and to improve the recognition performance in such a way as to compare them. Using this method, it was confirmed that the performance is improved by about 1.3% compared to the case of ignoring the occlusion situation and performing the recognition. Although the figure of about 1.3% is not absolutely high, it is not something that can be ignored in an environment where performance is somewhat converged, such as identity recognition. Instead of creating a new recognition model, it has the advantage of being able to improve performance while utilizing existing models. Besides, it is determined that the study of this direction is necessary because the perceived performance can be more helpful.

Jaeyoon Jang received the PhD degree in ICT from University of Science and Technology (UST) in 2019. He has been a research scientist at ETRI since 2019. His research interests are HMI and face recognition.

Hosub Yoon received the BS and MS degrees in computer science from SoongSil University, Seoul, Korea, in 1989 and 1991. He received the PhD degree in image processing from KAIST, Daejeon, in 2003. He joined KIST/SERI in 1991 and transferred to ETRI in 1999. He is a project leader of HMI(Human Machine Interaction) group in intelligence robotics division. His major research interests include HRI, AI, image processing, deep learning, audio processing and pattern recognition.

Jaehong Kim Received his PhD degree in computer engineering from Kyungpook National University, Daegu, Rep. of Korea, in 2006. He has been a research scientist at ETRI since 2001. His research interests include socially assistive robotics for the elderly, human-robot interaction, and social HRI frameworks

978-1-7281-6211-9/20/$31.00 ©2020 IEEE

II. RELATES WORKS

Research in the field of detecting and recognizing user face information has been steadily developed in various directions. In the case of face detection, starting from a representative study of the early deep learning [1, 2, 3], recent studies have come out detectable studies for extreme posture and size [4, 5, 6, 7, 8, 9]. So in the past, face detection itself often failed when information loss was caused by occlusion. In recent years, however, it has been possible to detect the face without being affected by accessories, obscuring, etc.

Landmark detection was mainly used in the early days by simply inferring the position in the image [10, 11, 12, 13], but recent studies have proposed to improve accuracy by combining 3D modeling [14, 15, 16, 17, 18]. So, as with face detection, it is possible to find the location of the main feature points in the face even if there is some information loss.

In addition to detection, various studies have been proposed in terms of identity recognition. Beyond the identity recognition performed in the ideal environment, a robust recognizer is being proposed in more and more different types, backgrounds, lighting, etc. [19, 20, 21, 22]. In addition to improving overall recognition performance, research on robust recognizers continues even when some information is lost due to occlusion [23, 24, 25, 26].

III. EXPERIMENTS & RESULT

This chapter deals with the influence of identity recognition on each area of the face and how to determine whether the face is covered. And it refers to a method for enhancing the recognition performance in the situation of obscuring through a method of extracting the feature using only the area that can extract the characteristic feature vector well.

A. *Mask detection*

In previous studies, it was common to have problems in the face detection phase when facial occlusion occurs due to factors such as masks. However, recent studies have shown that the performance of the detector has improved significantly, making it possible to detect the face and landmark with little to no effect by obscuring it. To find masks, we detected the face region and generated normalized face images using a landmark. Then remove the remaining area except for the candidate area of the mask in the normalized face area and generated a classification model using only the area around the mouth. The method for normalizing the face area has been applied the same way as proposed in [28]. Then, the value of the top 50% of the normalized face was replaced with zero, so that it was not used for calculation. The form of training image used for the detection mask is shown in figure 2.

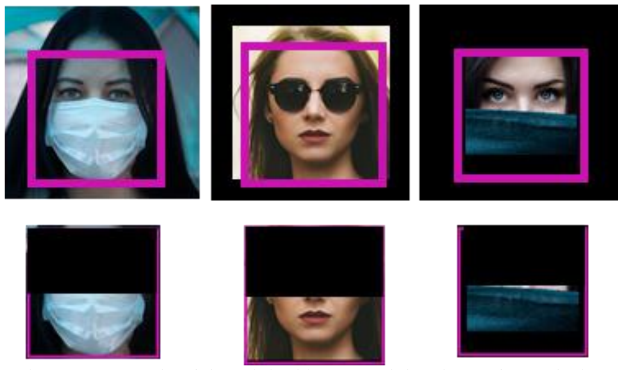

Figure 2. Example of the original image and the picture for occlusion detection

When training classifiers, we added branches to the existing model to take advantage of features extracted from the identity recognition model without learning a separate new classifier, and we designed it by adding a layer. There are two advantages to using this method. One is that the well-learned facial feature extractor can be recycled. Because it will perform recognition in the same domain, it works well even if it is used without an update. The other is that we can save memory by sharing the weight parameters. These tricks are not important in a high-end environment, but they require consideration where resources are limited, such as in a robot environment. This designed architecture is shown in the below figure 3.

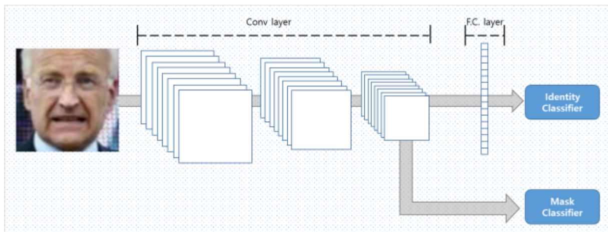

Figure 3. Identity recognition network & mask classifier

The data used in the training used the images collected on the web and images that artificially overlay the virtual mask on the Multi-PIE database [30]. When synthesizing, the mask image was warped to fit the landmark and overlaid. Because it is so clearly a matter of separating different images, it was possible to correctly classify all of the data. However, in the case of wild environments, it is necessary to supplement the data to improve the performance of the classifier because it includes a variety of changes that are not included in the training.

B. *Identification using information lost image*

It is common to assume that the environment can generally identify registered ideally controlled, registering the front face without face occlusion. However, because the actual perception situation is an uncontrolled environment, various changes can be reflected. So we conducted an experiment to determine the difference in the recognition performance of the case that part of the face is obscured by a mask. For performance evaluation, the virtual mask image was synthesized on a single face in lfw benchmark. Mask images used to generate virtual data were commonly used for disposable mask images. The synthetic example is as shown below figure 4.

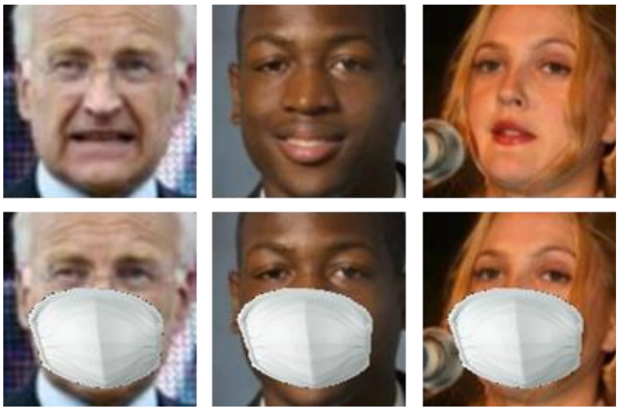

Figure 4. Artificially generated mask images.

Only one image of the two images for comparison was synthesized to the virtual mask image and performed the same lfw benchmark. In other words, the similarity between the face wearing the mask and the face without wearing is compared. As a result, it was obtained a recognition accuracy of about 96.6%, which is a similar result to the case of about 2-3% information loss occurred in the experiment for the preceding information loss.

C. *Proposed method*

We propose a system that combines the results of advanced experiments so that we can actively respond to situations of face occlusion. First, when the face is detected for identity recognition, the normalization process proceeds equally. Detects the face and landmark, corrects the rotation, and resize to a constant size to generate an input candidate image. Then copy the images to create a new candidate image. At this point, the new image is modified to the image that intentionally caused the loss of information by removing the bottom 50%. Then, the operation is performed by putting the modified network like figure 5 to calculate whether to wear a mask, the feature vector of the perfect image, and the feature vector for the occluded face. Finally, when comparing similarity with stored DB, similarity comparisons are performed using the feature vector calculated from the occluded face if occlusion occurs depending on whether the mask is worn or not, and if there is no obscure, similarity comparison is performed using the feature vector calculated from the perfect face. The entire system flow is the same as figure 5.

In order to determine the performance of the proposed system was applied equally to the experiments performed in the preceding C chapter. From the data of combining virtual masks on one face, it is determined whether or not the mask is present, and the features are extracted using only the areas that are meaningful if there is a mask. The result of applying this method is shown in table 1.

As you can see from the results above, we can see a slight improvement in performance compared to comparing images with masks as they are. In addition, there is no overhead in terms of computation time. It is possible to compare at almost the same time by using a network that is almost similar to the existing recognizer, and by computing the input image into a batch. The width of the performance improvement is not numerically a very large value, but it can be helpful to users in online testing.

Table 1. Performance comparison results of the proposed method

Method	Recognition Accuracy	performance decline ratio	synthesize a mask
Default recognizer	99.50%	0.00%	X
Direct compare	96.60%	-2.90%	0
Proposed method	**97.93%**	**-1.57%**	0

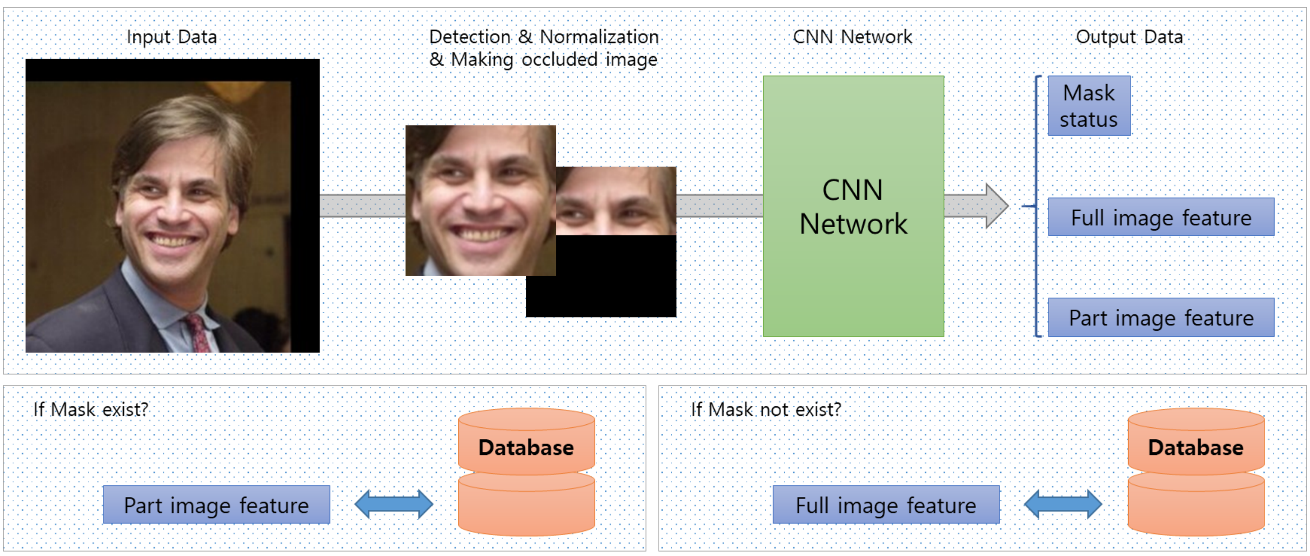

Figure 5. System flow

IV. CONCLUSION

In this paper, we propose a study on ways to proactively improve the recognition performance than if there is a loss of face information. First, the intermediate features of the identity recognizer were used to determine whether or not the appearance of obscures on the face. Then, we propose a method to improve recognition performance in a wild environment by selecting and using more suitable features. The advantage of the proposed method is that identity recognition can be performed using only actual information regardless of the type of occlusion using only meaningful information. Although the level of improvement in performance is not dramatic, it is a big advantage that it is possible to improve recognition performance in a particular environment without any cost. However, there are still some improvements. The first is that if an error occurs in the obscured judgment of the preceding step, the same result as the normal recognizer will be achieved. However, this is not a weakness because, in the worst case, the same results as the existing recognizer are achieved. The other is that there is a possibility that more information than necessary to be lost because it determines only the presence or absence of obscuring, and if the screening occurs, replacing the information of the fixed area to zero and calculating. In order to solve this problem, it is necessary to accurately detect the starting position of the occlusion or segmentation of the obscured area. Using such a method, it is expected that the identity characteristics will be extracted more efficiently and the better identity recognition results will be obtained by comparing only meaningful parts without loss of necessary information.

ACKNOWLEDGMENT

This work was supported by the IT R&D program of MOTIE/KEIT [10077553], Development of Social Robot Intelligence for Social Human-Robot Interaction of Service Robots.

REFERENCES

[1] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.

[2] Redmon, Joseph, and Ali Farhadi. "YOLO9000: better, faster, stronger." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[3] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.

[4] He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.

[5] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.

[6] Zhang, Jialiang, et al. "Feature agglomeration networks for single stage face detection." Neurocomputing 380 (2020): 180-189.

[7] Zhang, Zhishuai, et al. "Robust face detection via learning small faces on hard images." The IEEE Winter Conference on Applications of Computer Vision. 2020.

[8] Zhang, Kaipeng, et al. "Joint face detection and alignment using multitask cascaded convolutional networks." IEEE Signal Processing Letters 23.10 (2016): 1499-1503.

[9] Tan, Mingxing, Ruoming Pang, and Quoc V. Le. "Efficientdet: Scalable and efficient object detection." arXiv preprint arXiv:1911.09070 (2019).

[10] Qian, Shengju, et al. "Aggregation via Separation: Boosting Facial Landmark Detector with Semi-Supervised Style Translation." Proceedings of the IEEE International Conference on Computer Vision. 2019.

[11] Wang, Xinyao, Liefeng Bo, and Li Fuxin. "Adaptive wing loss for robust face alignment via heatmap regression." Proceedings of the IEEE International Conference on Computer Vision. 2019.

[12] Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Deep convolutional network cascade for facial point detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2013.

[13] Zhou, Erjin, et al. "Extensive facial landmark localization with coarse-to-fine convolutional network cascade." Proceedings of the IEEE international conference on computer vision workshops. 2013.

[14] Paulsen, Rasmus R., et al. "Multi-view consensus CNN for 3D facial landmark placement." Asian Conference on Computer Vision. Springer, Cham, 2018.

[15] Zhu, Xiangyu, et al. "Face alignment across large poses: A 3d solution." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[16] Wang, Yue, and Justin M. Solomon. "PRNet: Self-Supervised Learning for Partial-to-Partial Registration." Advances in Neural Information Processing Systems. 2019.

[17] Jourabloo, Amin, and Xiaoming Liu. "Large-pose face alignment via CNN-based dense 3D model fitting." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[18] Wang, Kai, et al. "A coarse-to-fine approach for 3D facial landmarking by using deep feature fusion." Symmetry 10.8 (2018): 308.

[19] Wen, Yandong, et al. "A discriminative feature learning approach for deep face recognition." European conference on computer vision. Springer, Cham, 2016.

[20] Liu, Weiyang, et al. "Sphereface: Deep hypersphere embedding for face recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[21] Wang, Hao, et al. "Cosface: Large margin cosine loss for deep face recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[22] Deng, Jiankang, et al. "Arcface: Additive angular margin loss for deep face recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

[23] Lu, Boyu, et al. "An experimental evaluation of covariates effects on unconstrained face verification." IEEE Transactions on Biometrics, Behavior, and Identity Science 1.1 (2019): 42-55.

[24] Mao, Li, Fusheng Sheng, and Tao Zhang. "Face Occlusion Recognition With Deep Learning in Security Framework for the IoT." IEEE Access 7 (2019): 174531-174540.

[25] Yu, Guanwen, and Zhaogong Zhang. "Face and occlusion Recognition Algorithm based on Global and Local." Journal of Physics: Conference Series. Vol. 1453. 2020.

[26] Zhao, Yunfei, et al. "Face Verification Between ID Document Photos and Partial Occluded Spot Photos." International Conference on Image and Graphics. Springer, Cham, 2019.

[27] Huang, Gary B., et al. "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments." 2008.

[28] Wu, Xiang, et al. "A light cnn for deep face representation with noisy labels." IEEE Transactions on Information Forensics and Security 13.11 (2018): 2884-2896.

[29] Jang, Jaeyoon, et al. "Robust deep age estimation method using artificially generated image set." ETRI Journal 39.5 (2017): 643-651.

[30] Gross, Ralph, et al. "Multi-pie." Image and Vision Computing 28.5 (2010): 807-813.