

Article

Improvement of Identity Recognition with Occlusion Detection-Based Feature Selection

Jaeyoon Jang *, Hosub Yoon and Jaehong Kim

Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, Korea; yoonhs@etri.re.kr (H.Y.); jhkim504@etri.re.kr (J.K.)

* Correspondence: jangjy@etri.re.kr

Abstract: Image-based facial identity recognition has become a technology that is now used in many applications. This is because it is possible to use only a camera without the need for any other device. Besides, due to the advantage of contactless technology, it is one of the most popular certifications. However, a common recognition system is not possible if some of the face information is lost due to the user's posture or the wearing of masks, as caused by the recent prevalent disease. In some platforms, although performance is improved through incremental updates, it is still inconvenient and inaccurate. In this paper, we propose a method to respond more actively to these situations. First, we determine whether an 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, the results of reducing the recognition performance drop in certain situations were confirmed. Using this technique, we confirmed a performance improvement of about 1~3% in a situation where some information is lost. Although the performance is not dramatically improved, it has the big advantage that it can improve recognition performance by utilizing existing systems.

Keywords: face verification; occlusion detection; feature selection



Citation: Jang, J.; Yoon, H.; Kim, J. Improvement of Identity Recognition with Occlusion Detection-Based Feature Selection. *Electronics* **2021**, *10*, 167. <https://doi.org/10.3390/electronics10020167>

Received: 9 December 2020

Accepted: 10 January 2021

Published: 13 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. 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 settlements 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 are still problems that remain. The biggest problem among them is that the entire face needs to be seen correctly for accurate certification. If wearing a mask due to cold weather or for protection against diseases such as COVID-19, it is impossible to be recognized as 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 if even part of the face area is cut off due to the user's posture during interaction with the robot.

For various reasons, as shown in Figure 1, facial information is lost, but we nevertheless want to gain a strong recognition performance. Common studies so far have expected a single recognition model to be able to develop in a variety of environments and respond to all exceptional situations. Of course, due to the development of deep learning technology, if only enough databases for various environments were built, we could expect such performance, but realistically, it is impossible to gather data about all possible situations. Therefore, we conducted a study on how to recognize identity more robustly in situations such as those involving use of masks and screen truncation, re-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. This is because follow-up experiments can

be designed only after checking the effect of the degree of face covering on recognition performance.

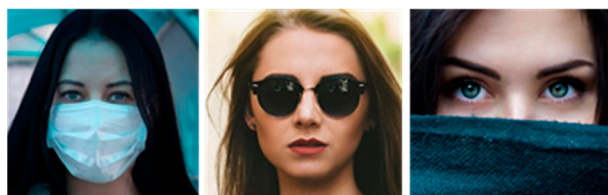


Figure 1. Cases of occlusion on the face due to various factors.

In addition, we experimented with identity recognizers that could respond to situations where faces were obscured. To respond to situations where the face was obscured, we first recognized whether the face is obscured and then designed a method for calculating the feature vectors for when the face is obscured and when it is not obscured, respectively. When the cover occurs, the feature vector is calculated only in the unobstructed area to minimize deformation due to external factors. When comparing identity similarity, a database of feature vectors with obscurity was separately managed to prevent situations in which similarities between feature vectors with different characteristics were compared. 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.

The proposed method also shows a decline in performance compared to the intact situation, but it has the advantage that the degree of performance decline can be greatly alleviated. A typical identity recognition model is designed to respond to changes in posture, lighting, etc., but accurate identity features can be calculated only when all face regions are visible. However, the proposed method has the advantage of suppressing the decline in identity recognition performance in a situation where loss of information occurs by determining whether or not it is obscured and helping to select an identity feature suitable for the situation. Our study also has the advantage of being able to recycle existing recognition models by including a simple add-on instead of creating a new recognition model. The recognition model corresponding to the obscured environment recently studied is a newly trained recognition model using a new method. In this case, the identity recognition performance of the existing ideal situation may be affected, and a lot of resources are consumed to train a new model. However, the proposed method has an advantage in that it uses few additional resources, does not affect the existing identity recognition performance at all, and can suppress performance degradation in special situations.

2. Related Work

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 early deep learning [1–3], recent studies have come out with detectable studies for extreme posture and size [4–9]. 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 of the landmark in an image [10–13], but recent studies have proposed to improve accuracy by combining 3D modeling [14–18]. Therefore, 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. In the past, human-designed and calculated identity vectors were mainly used. A personal feature vector was used to convert regional images such as LBP (Local Binary Pattern), MCT (Modified Census Transform), HoG (Histogram of Gradient orientations), etc. [19–21], or using a Gabor filter to imitate the visual information of the person [22]. Because these studies reflected the local characteristics of images in feature

vectors, there was a problem of significant performance degradation if there was an image transformation. Subsequently, deep learning techniques were introduced, and methods for extracting high-dimensional identity features from images began to be proposed [23–26]. These early studies of deep learning aimed at calculating feature vectors that distinguish each identity as well as possible by analyzing images from the training dataset as they are. Later, a study was conducted to maximize the distance between classes in the feature space using margin loss or angular loss [27–29]. Although these attempts were enhanced in adaptability to factors such as posture, facial expressions, and the environment, there were still limitations in the wild environment. However, as the development of deep learning technology has progressed, beyond identity recognition performed in the ideal environment, a robust recognizer is being proposed in more and more different types, backgrounds, lighting, etc. [30–34]. Instead of simply learning a network that categorizes the training dataset well, it creates a virtual dataset to learn the external environment or combines multiple networks into an ensemble to improve performance and design a multi-task network to make various attempts in which information complementarily aids learning. In addition to improving overall recognition performance, research on robust recognizers continues even when some information is lost due to occlusion [35–39]. These studies are developing an identity identification method that corresponds to the case where there is a loss of information due to obscuration, etc., rather than cases where the face is seen completely. However, this new study has a limitation in that it not only affects recognition performance in a general environment but also needs to train a new recognition network.

3. Experiments and Results

This chapter deals with the influence of identity recognition on each area of the face and how to determine whether the face is covered. It also introduces the method of determining obscurity using facial feature vectors of existing recognizers. Finally, if obscuring occurs in the face area, we suggest a trick to calculate the identity feature vector using only the area where the identity feature vector can be well calculated.

3.1. Influence on Occlusion

As previously mentioned, occlusion of the face can be caused by a variety of situations. General studies so far have not been approached in a way that directly responds to situations such as occlusion. Regardless of what form of face comes in, studies have been conducted in the direction of extracting the characteristic feature vector of the individual robustly. We decided that if we manage these cases separately, it will be helpful for recognition, and we checked the performance differences according to the degree of occlusion of the face area. In other words, it is an experiment to find the face area that is observed to be important in identity recognition. For experimental data, the verification benchmark of the LFW (Labeled Faces in the Wild) [40] data is used. As we gradually occluded the image from the complete image, we confirmed the change in the recognition rate. Because it was an experiment to confirm the difference in recognition performance due to information loss, both images used for verification were partially covered at the same rate. The result was used for a value of 1.0 Equal Error Rate (EER) according to the benchmark protocol. An example of the image used in the experiment is shown in Figure 2.

In this study, an experiment for horizontal occlusion, which generally occurs easily when using an accessory, was conducted preferentially. The range of occlusions was set from 0% to 60% of the top/bottom area, the pixel value was modified to 0, and the information was intentionally lost. For comparison of influences, the occlusion range was changed in 1% increments, and the difference was confirmed. To compare the difference in the recognition performance according to the amount of information on the face, both images were applied to the same information loss. For performance comparison, the identity feature extraction network used lightCNN [41] architecture. When training a feature extractor, we applied the accessories augmentation method proposed in [42]. The generated recognition model has a recognition performance of 99.58% when there is no

information loss. The difference in recognition performance due to the loss of information for each rate is shown in Figure 3.

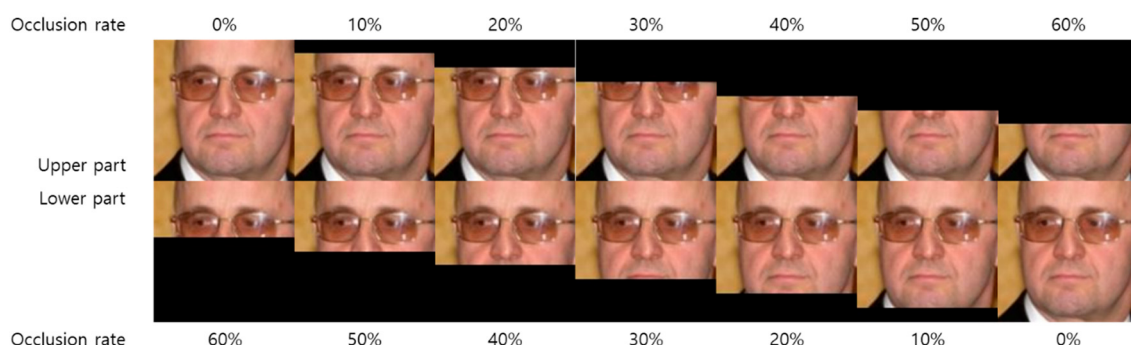


Figure 2. Example of occlusion images. The image is occluded in a certain proportion in up/down directions.

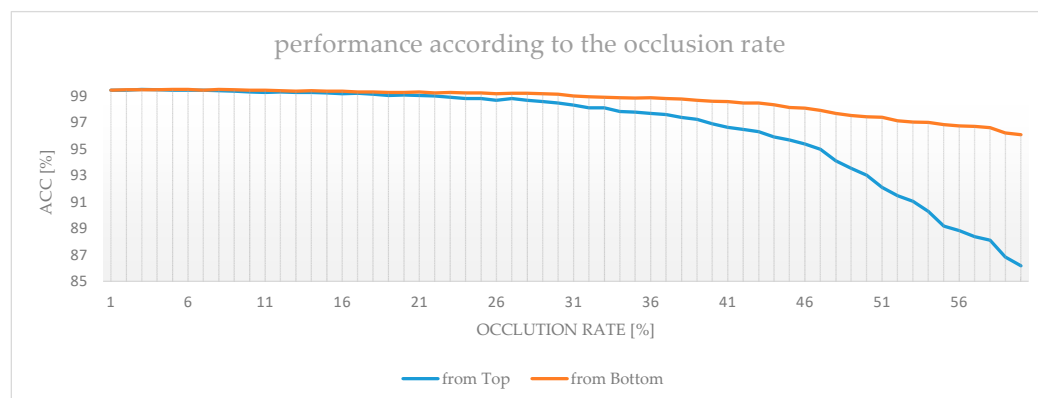


Figure 3. Changes in verification performance according to the occlusion rate.

The results of the experiment showed that the more information was lost, the lower the overall recognition performance. In particular, it was possible to confirm that the upper portion of the face had a greater effect on identification. If the upper part of the face, including the eyes, was occluded, we could confirm that identity recognition performance was critically declining. In addition, if the bottom part of the face was obscured and the loss of information occurred, it was possible to see a performance drop of about 4% compared to the case of the entire face being visible. However, as opposed to the lower part of the face, it means that even without the lower part of the face, it can achieve an accuracy rate of 96% of the maximum performance. However, this change causes more performance decline than the numerical value in the online test, so a solution is needed.

3.2. Identity Recognition Using Partially Obscured Images

In a general identity recognition environment, it is assumed that the registration environment is controllable, so no occlusion occurs and the front face is registered. However, because the actual perception situation is an uncontrolled environment, various changes can be reflected. Therefore, we experimented to determine the differences in the recognition performance in the case that part of the face is obscured by a mask. For performance evaluation, a virtual mask image was synthesized on a single face in the LFW benchmark. Mask images used to generate virtual data were commonly used for disposable mask images. In addition, the virtual mask patch presented in the NIST (National Institute of Standards and Technology) report [43] on the effect of masks on identity recognition was applied with the same rule. The synthetic example is shown in Figure 4.

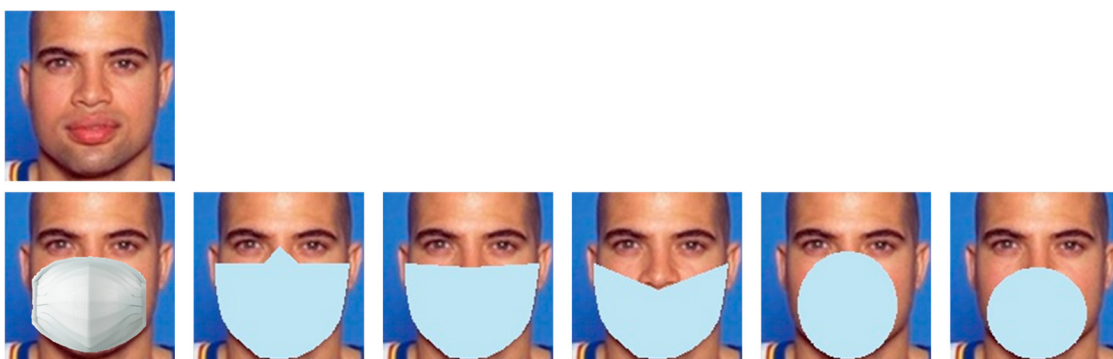


Figure 4. Artificially generated mask images.

Only one image of the two images for comparison was synthesized as the virtual mask image and compared with the same LFW benchmark. In other words, the similarity between the face wearing the mask and the face without it is compared. As a result, a recognition accuracy of about 94–97% was obtained, which is a similar result to the case of about 2–5% of information loss that occurred in the experiment for the preceding information loss. The performance decline of 2~5% is the same level as the degenerating of the level of identification technology for 6~8 years, so it is fatal to the identification recognition using the face image. The specific performance of each case can be checked in the result table in Section 3.3.2.

3.3. Proposed Method and Result

In this section, we propose a method to improve identity recognition performance in special situations based on the previous experimental results. We propose a method of determining whether to wear a mask by recycling the existing identity feature vector and a method of minimizing the influence of the mask. Finally, we confirm that the proposed add-on can be used to improve identity recognition performance in a masked situation.

3.3.1. Mask Detection with Recycled Identity Features

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 detectors 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, we removed 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 in the same way as proposed in [32]. Then, the value of the top 40~50% 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 5.



Figure 5. 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 training a separate new classifier, and we designed it by adding a layer. The weight of the existing recognition

model is then used only as a feature extractor without updating it. That way, the mask detection layer is trained in the form of a classifier that classifies whether or not obscuring has occurred from identity features. Because it is a simple form with two-class classification, the loss is defined and used as follows:

$$\text{loss}_{\text{ce}} = - \sum_{i=1}^2 t_i \log(s_i) = -t_1 \log(s_1) - (1 - t_1) \log(1 - s_1) \quad (1)$$

There are two advantages to using this method. One is that the well-trained 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 6.

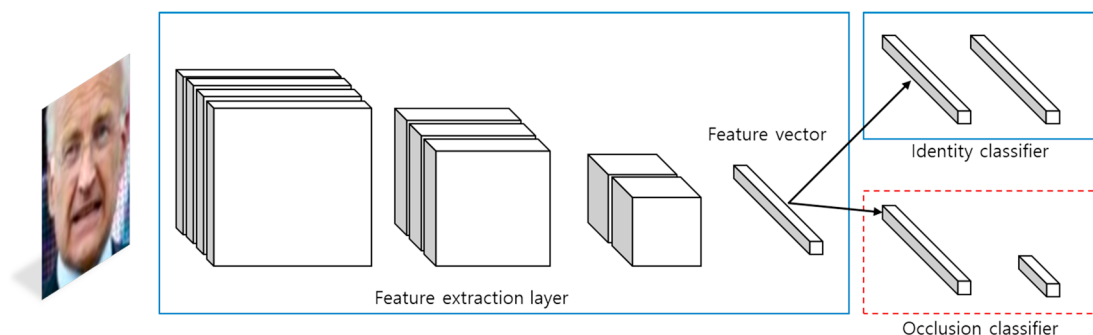


Figure 6. Identity recognition network and occlusion classifier. The blue solid line represents the existing identity recognizer, and the red dotted line represents the added occlusion classifier.

The data used in the training were images collected on the web and images that artificially overlay a virtual mask on the Multi-PIE database [44]. When synthesizing, the mask image was warped to fit the landmark and overlaid. One of 346 subjects on the Multi-PIE database was separated into a validation set by the cross-subject method, and the remaining identities were used for learning, and all lighting environments and posture environments were reflected. For the mask classifier, about 200,000 pieces of training data were used and about 700 pieces of validation data were used. 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.

3.3.2. Feature Selection with Mask Detection

We propose a system that combines the results of advanced experiments so that we can actively respond to situations of face occlusion. This is a method for calculating a feature with only necessary information by determining a meaningful information region within the face region and preventing the remaining regions from being involved in identification feature extraction. The proposed method is as follows. First, when the face is detected for identity recognition, the normalization process proceeds equally. It detects the face region and landmark, corrects the rotation, and resizes to a constant size to generate an input candidate image. Then, the image is copied to create a new candidate image. New candidate images are created for two purposes. One is to recognize whether occlusion such as that by a mask has occurred, and the other is to use only a significant area for identification feature calculation. In this step, the new candidate image is modified to the image that intentionally caused the loss of information by padding 50–60% with zeros. Then, the operation is performed by using the modified network such as that in Figure 6 to

calculate whether the person is wearing a mask, the feature vector of the perfect image, and the feature vector for the occluded face. Finally, when comparing the similarity with the stored database, similarity comparisons are performed using the feature vector calculated from the occluded face if occlusion occurs depending on whether a mask is worn or not, and if there is no obscuring, the similarity comparison is performed using the feature vector calculated from the perfect face. The entire system flow is the same as in Figure 7.

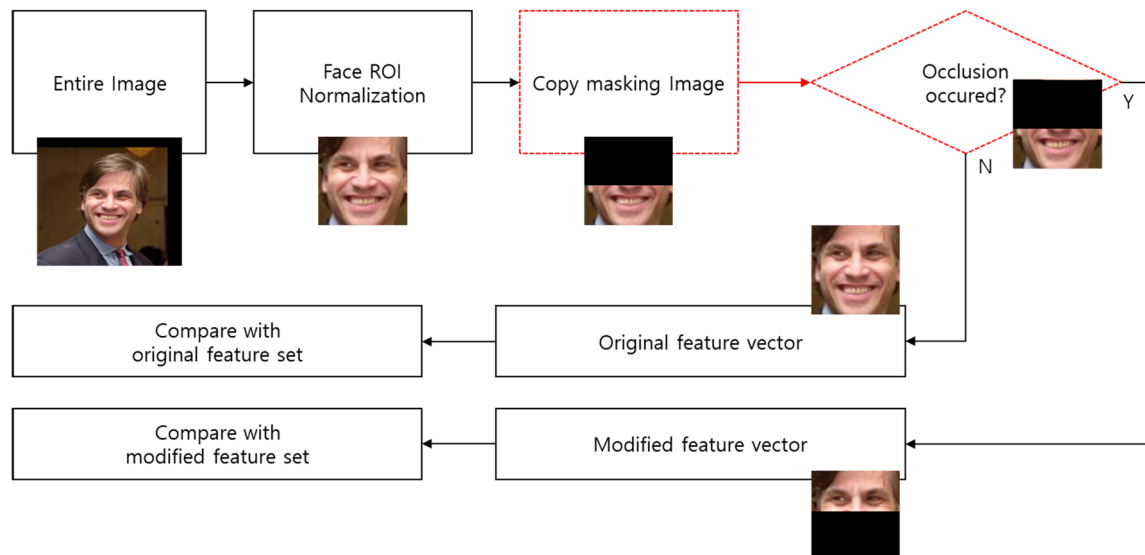


Figure 7. The entire system flow. The red dotted line is the operation unit added to the existing recognition system.

The difference between the proposed method and the existing identity recognition system is the red dotted line in Figure 7. Two steps have been added in the identity recognition process, but this process has little effect on the overall computation. Since the process of creating a masking image does not require complex image processing, only a process of padding a value of zero for a part of the area causes a small additional use of resources. The part that determines occlusion is also combined with the existing network as shown in Figure 6, so the identity feature vector and occlusion recognition result can be obtained with a single feed-forward operation. In this step, three types of candidate images are used, but if the three images are operated in mini-batch, the result can be obtained in almost the same operation time as the existing procedure. In the process of registering a new identity, to use the proposed method, the identity features calculated from the original image and the identity features calculated from the modified image are stored respectively. In contrast to the existing method, the only overhead of the proposed method is that the gallery dataset has to be stored separately. However, since the resources required to store the identity feature vector are not large, the burden is light relative to the performance improvement.

We performed three experiments to verify the performance of the proposed method. These include feature vector visualization, influence on virtual masks, and performance verification in other identity recognizers. First, we visualized the proposed part image feature through t-SNE to confirm whether it maintains the appropriate distance from the existing full image feature. The image from the Multi-PIE dataset was used for visualization. The feature vector was calculated from the image that synthesized the original image with a virtual mask, and the image to which the add-on was applied was projected in the 2D plane. The visualized feature plane is shown in Figure 8. As a result of the visualization, it was confirmed that the distance of the full image feature between the face without the mask and the image with the mask was far greater than the distance of the partial image feature to which the add-on was applied. In addition, it was confirmed that the partial

image feature, which applied the add-on, is not close enough to be implicitly distinguished from other classes.

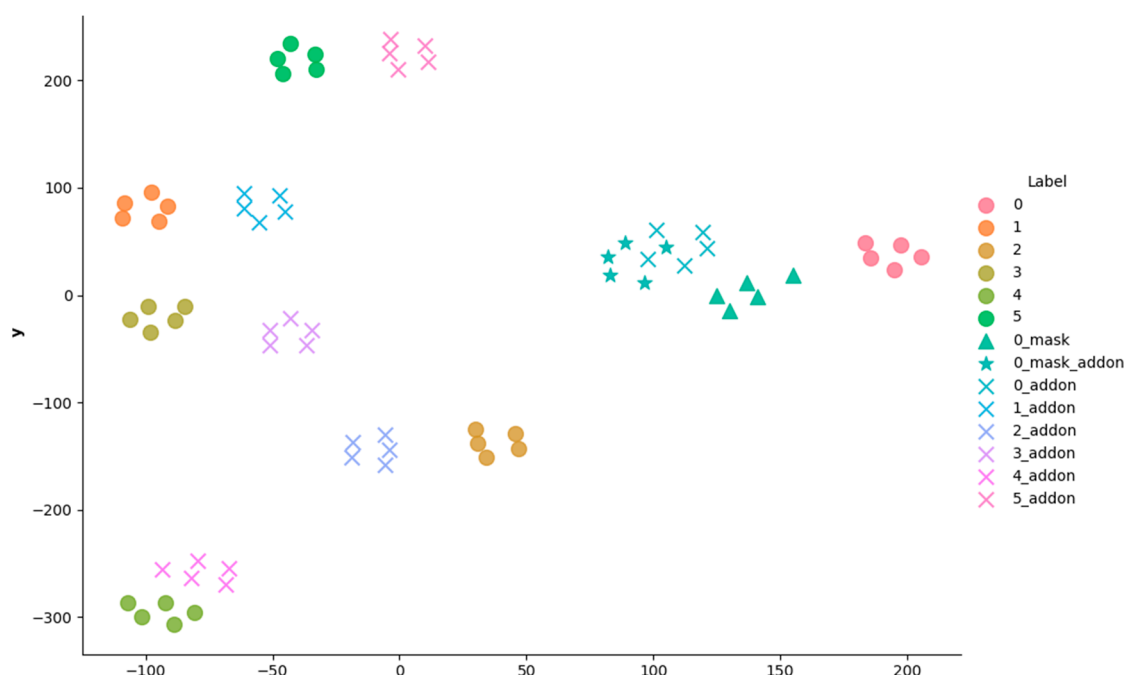


Figure 8. Visualized feature plane. In the label, the number means the class, “_mask” means the composite of the virtual mask, and “_addon” refers to a partial image feature that applies the add-on.

Second, to compare the distance between the vectors projected in dimensions and to confirm the actual verification performance, the performance was evaluated by applying the method used in Section 3.2 to the benchmark database. A virtual mask was synthesized in the image of one of the verification pairs, and the verification performance was calculated when the add-on was applied and otherwise. The types of virtual masks used in the experiment are shown in Figure 4 and were implemented to automatically be generated according to the detected landmark. The identity recognizer used for performance evaluation was not an open recognition model but a self-developed recognizer with 99.58% performance in the LFW benchmark database. The result of applying this method is shown in Figure 9. First, in an environment where a virtual mask was not added, it was confirmed that information loss occurs when the add-on is added, so the recognition performance decreases. In the case of adding a virtual mask, recognition performance decreases in all cases, but in most cases, it can be confirmed that the recognition performance recovers when the add-on is added. We confirmed that if more than 60% of the entire image is corrected, the amount of information loss increases and the recognition performance is relatively lowered. Through this experiment, the performance improvement was the biggest when the add-on was applied with an image that was modified at a rate of 55% on average, although there were some variations.

Finally, we performed experiments to confirm if the same effect can be obtained in other identification recognizers. For comparison, we used MobileFace [45] and FaceNet (InceptionResnetV1) [46] as recognizers. The virtual mask was used with the same mask template of NIST as the previous experiment and the add-on was applied with the modified image at a 55% ratio. The overall performance is shown in Table 1.

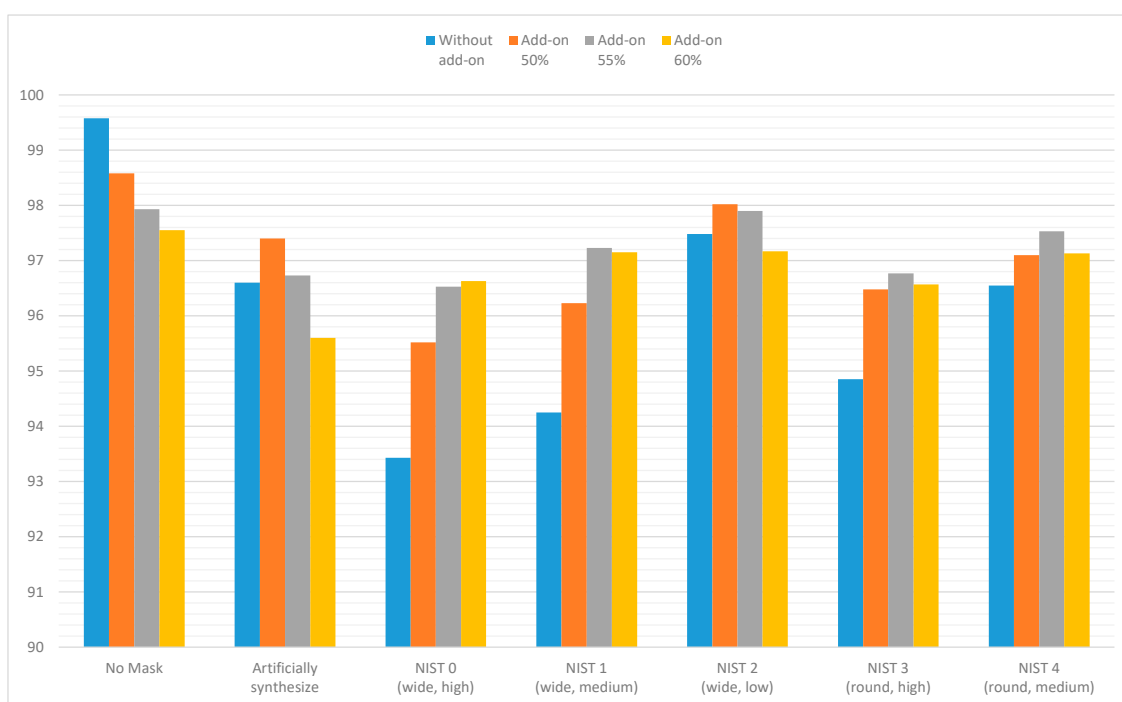


Figure 9. Performance graphs according to test cases.

Table 1. Performance comparison results of the proposed method.

Recognizer	Method	NIST Mask 0	NIST Mask 1	NIST Mask 2	NIST Mask 3	NIST Mask 4	w/o Mask
MobileFace	+mask	88.91	90.89	96.75	90.79	95.33	99.18
	+mask						
	+add-on	95.34	96.89	96.86	94.99	97.06	
Inception Resnet-V1	+mask	87.17	89.69	96.78	87.24	91.10	99.05
	+mask						
	+add-on	95.15	96.59	96.92	94.32	96.90	
Ours	+mask	93.43	94.25	97.48	94.85	96.55	99.58
	+mask						
	+add-on	96.53	97.23	97.90	96.77	97.53	

Similar to the previous experiment, when a virtual mask was added, the recognition performance was reduced by around 10%. In some experiments, we confirmed that the recognition performance dropped by even more, which is presumed to be due to the lower performance of the identity recognizer itself which is more affected. However, it was confirmed that a large part of the lost performance was restored when the add-on was applied.

As the results of the experiment show, we can see a slight performance improvement compared to images with masks as they are. The performance when using a mask decreased by up to 5% depending on the type of mask, but it was confirmed that the drop rate can be suppressed by 50% when the add-on is applied. In addition, there is no overhead in terms of computation time. It is possible to compare images in almost the same time by using a network that is similar to the existing recognizer and by computing the input image into a mini-batch. The rate of the performance improvement is not numerically a very large value, but this method maintains existing recognition performance in an environment where a mask is not worn, and when a mask is worn, it is helpful in the actual recognition environment by suppressing performance declines.

4. Conclusions

In this paper, we propose a solution to the problem where the performance of an existing identity recognizer is degraded in the situation where occlusion occurs. A summary of the methods for suppressing degradation in certain situations without compromising the performance of existing identity recognizers is as follows. First, we included an add-on that can recognize whether occlusion has occurred using the existing identity feature vector. Then, according to the recognition result of the add-on, we compared the similarity by selecting an appropriate vector among the existing identity feature vectors and the identity feature vectors that minimize the influence of the occlusion. 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 rate of improvement in performance is not dramatic, it is a big advantage in that it is possible to improve recognition performance in a particular environment without any cost. However, there are still some improvements to be made. Firstly, 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 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 identity characteristics will be extracted more efficiently and better identity recognition results will be obtained by comparing only meaningful parts without loss of necessary information.

Author Contributions: Conceptualization, J.J., H.Y. and J.K.; methodology, J.J.; software, J.J.; validation, J.J.; formal analysis, J.J. and H.Y.; investigation, J.J. and H.Y.; resources, J.J., H.Y. and J.K.; data curation, J.J. and J.K.; writing—original draft preparation, J.J.; writing—review and editing, J.J., H.Y. and J.K.; visualization, J.J.; supervision, H.Y. and J.K.; project administration, H.Y.; funding acquisition, H.Y. and J.K. All authors have read and agreed to the published version of the manuscript.

Funding: 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, and the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant, funded by the Korea government (MSIP) (No. 2020-0-00842, Development of Cloud Robot Intelligence for Continual Adaptation to User Reactions in Real Service Environments).

Data Availability Statement: Data available in a publicly accessible repository that does not issue DOIs Publicly available datasets were analyzed in this study. This data can be found here: [<http://vis-www.cs.umass.edu/lfw/>, <http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html>].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. GIRSHICK, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 11–18 December 2015; pp. 1440–1448.
2. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
3. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
4. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
5. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
6. Zhang, J.; Wu, X.; Hoi, S.C.; Zhu, J. Feature agglomeration networks for single stage face detection. *Neurocomputing* **2020**, *380*, 180–189. [[CrossRef](#)]

7. Zhang, Z.; Shen, W.; Qiao, S.; Wang, Y.; Wang, B.; Yuille, A. Robust face detection via learning small faces on hard images. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 1361–1370.
8. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [[CrossRef](#)]
9. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
10. Qian, S.; Sun, K.; Wu, W.; Qian, C.; Jia, J. Aggregation via separation: Boosting facial landmark detector with semi-supervised style translation. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 10153–10163.
11. Wang, X.; Bo, L.; Fuxin, L. Adaptive wing loss for robust face alignment via heatmap regression. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 6971–6981.
12. Sun, Y.; Wang, X.; Tang, X. Deep convolutional network cascade for facial point detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 3476–3483.
13. Zhou, E.; Fan, H.; Cao, Z.; Jiang, Y.; Yin, Q. Extensive facial landmark localization with coarse-to-fine convolutional network cascade. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Sydney, Australia, 2–8 December 2013; pp. 386–391.
14. Paulsen, R.R.; Juhl, K.A.; Haspang, T.M.; Hansen, T.; Ganz, M.; Einarsson, G. Multi-view consensus CNN for 3D facial landmark placement. In *Asian Conference on Computer Vision*; Springer: Cham, Germany, 2018; pp. 706–719.
15. Zhu, X.; Lei, Z.; Liu, X.; Shi, H.; Li, S.Z. Face alignment across large poses: A 3d solution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 146–155.
16. Wang, Y.; Solomon, J.M. PRNet: Self-supervised learning for partial-to-partial registration. In Proceedings of the Advances Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 8814–8826.
17. Jourabloo, A.; Liu, X. Large-pose face alignment via CNN-based dense 3D model fitting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4188–4196.
18. Wang, K.; Zhao, X.; Gao, W.; Zou, J. A coarse-to-fine approach for 3D facial landmarking by using deep feature fusion. *Symmetry* **2018**, *10*, 308. [[CrossRef](#)]
19. Ahonen, T.; Hadid, A.; Pietikäinen, M. Face recognition with local binary patterns. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 5–12 September 2004; pp. 469–481.
20. Rodriguez, Y.; Marcel, S. Face authentication using adapted local binary pattern histograms. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 321–332.
21. Déniz, O.; Bueno, G.; Salido, J.; De la Torre, F. Face recognition using histograms of oriented gradients. *Pattern Recognit. Lett.* **2011**, *32*, 1598–1603. [[CrossRef](#)]
22. Barbu, T. Gabor filter-based face recognition technique. *Rom. Acad.* **2010**, *11*, 277–283.
23. Taigman, Y.; Yang, M.; Ranzato, M.A.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708.
24. Sun, Y.; Wang, X.; Tang, X. Deep learning face representation from predicting 10,000 classes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1891–1898.
25. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. 2015. Available online: <https://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/parkhi15.pdf> (accessed on 9 December 2020).
26. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.
27. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. Spheraface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 June 2017; pp. 212–220.
28. Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Liu, W. Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5265–5274.
29. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 4690–4699.
30. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A discriminative feature learning approach for deep face recognition. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 499–515.
31. Morales, A.; Fierrez, J.; Vera-Rodriguez, R. SensitiveNets: Learning agnostic representations with application to face recognition. *arXiv* **2019**, arXiv:1902.00334.
32. Zhao, J.; Li, J.; Tu, X.; Zhao, F.; Xin, Y.; Xing, J.; Feng, J. Multi-prototype networks for unconstrained set-based face recognition. *arXiv* **2019**, arXiv:1902.04755.
33. Pearline, S.A. Face Recognition under Varying Blur, Illumination and Expression in an Unconstrained Environment. *arXiv* **2019**, arXiv:1902.10885.

34. Ming, Z.; Xia, J.; Luqman, M.M.; Burie, J.C.; Zhao, K. Dynamic Multi-Task Learning for Face Recognition with Facial Expression. *arXiv* **2019**, arXiv:1911.03281.
35. Lu, B.; Chen, J.C.; Castillo, C.D.; Chellappa, R. An experimental evaluation of covariates effects on unconstrained face verification. *IEEE Trans. Biom. Behav. Identity Sci.* **2019**, *1*, 42–55. [[CrossRef](#)]
36. Mao, L.; Sheng, F.; Zhang, T. Face Occlusion Recognition with Deep Learning in Security Framework for the IoT. *IEEE Access* **2019**, *7*, 174531–174540. [[CrossRef](#)]
37. Yu, G.; Zhang, Z. Face and occlusion Recognition Algorithm based on Global and Local. *J. Phys. Conf. Ser.* **2020**, *1453*, 012019. [[CrossRef](#)]
38. Zhao, Y.; Wei, S.; Jiang, X.; Ruan, T.; Zhao, Y. Face Verification between ID Document Photos and Partial Occluded Spot Photos. In Proceedings of the International Conference on Image and Graphics, Beijing, China, 23–25 August 2019; pp. 94–105.
39. Duan, Q.; Zhang, L. BoostGAN for Occlusive Profile Face Frontalization and Recognition. *arXiv* **2019**, arXiv:1902.09782.
40. Huang, G.B.; Mattar, M.; Berg, T.; Learned-Miller, E. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*; Technical Report; University of Massachusetts: Amherst, MA, USA, 2007; pp. 7–49.
41. Wu, X.; He, R.; Sun, Z.; Tan, T. A light cnn for deep face representation with noisy labels. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2884–2896. [[CrossRef](#)]
42. Jang, J.; Jeon, S.H.; Kim, J.; Yoon, H. Robust deep age estimation method using artificially generated image set. *ETRI J.* **2017**, *39*, 643–651. [[CrossRef](#)]
43. Ngan, M.L.; Grother, P.J.; Hanaoka, K.K. Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face Recognition Accuracy with Masks Using pre-COVID-19 Algorithms. NIST Interagency/Internal Report (NISTIR)–8311. 2020. Available online: <https://www.nist.gov/publications/ongoing-face-recognition-vendor-test-frvt-part-6a-face-recognition-accuracy-masks-using> (accessed on 9 December 2020).
44. Gross, R.; Matthews, I.; Cohn, J.; Kanade, T.; Baker, S. Multi-pie. *Image Vis. Comput.* **2010**, *28*, 807–813. [[CrossRef](#)] [[PubMed](#)]
45. Chen, S.; Liu, Y.; Gao, X.; Han, Z. Mobilefacenet: Efficient cnns for accurate real-time face verification on mobile devices. In *Chinese Conference on Biometric Recognition*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 428–438.
46. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. *AAAI Conf. Artif. Intell.* **2017**, *31*, 1.