



Article ID 1007-1202(2023)02-0141-09

DOI <https://doi.org/10.1051/wujns/2023282141>

Periocular Biometric Recognition for Masked Faces

□ HUANG Qiaoyue, TANG Chaoying[†], ZHANG Tianshu

College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, Jiangsu, China

© Wuhan University 2023

Abstract: Since the outbreak of Coronavirus Disease 2019 (COVID-19), people are recommended to wear facial masks to limit the spread of the virus. Under the circumstances, traditional face recognition technologies cannot achieve satisfactory results. In this paper, we propose a face recognition algorithm that combines the traditional features and deep features of masked faces. For traditional features, we extract Local Binary Pattern (LBP), Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradient (HOG) features from the periocular region, and use the Support Vector Machines (SVM) classifier to perform personal identification. We also propose an improved Convolutional Neural Network (CNN) model Angular Visual Geometry Group Network (A-VGG) to learn deep features. Then we use the decision-level fusion to combine the four features. Comprehensive experiments were carried out on databases of real masked faces and simulated masked faces, including frontal and side faces taken at different angles. Images with motion blur were also tested to evaluate the robustness of the algorithm. Besides, the experiment of matching a masked face with the corresponding full face is accomplished. The experimental results show that the proposed algorithm has state-of-the-art performance in masked face recognition, and the periocular region has rich biological features and high discrimination.

Key words: masked face recognition; periocular; Visual Geometry Group (VGG); Local Binary Pattern (LBP); Scale-Invariant Feature Transform (SIFT); Histogram of Oriented Gradient (HOG); Support Vector Machines (SVM)

CLC number: TP 391.41

0 Introduction

Biometric recognition uses fingerprints, veins, faces, DNA, etc. for the verification and identification of personal identity^[1]. These features should be unique, ubiquitous, and invariant. The automatic identity authentication system based on biometrics such as fingerprints and faces has developed relatively maturely^[2]. However, under the COVID-19 pandemic situation, people usually wear masks in daily life, making conventional facial recognition technology inefficient. According to a prelimi-

nary study by the National Institute of Standards and Technology (NIST), even the best of the 89 commercial facial recognition algorithms tested had error rates between 5% and 50% in matching digitally applied face masks with photos of the same person without a mask^[3]. Fingerprints need to be collected by contact instruments, which is not conducive to epidemic prevention. These existing identity authentication systems have exposed certain drawbacks during the epidemic.

Therefore, biometric recognition for facial regions above the mask has become an important and novel re-

Received date: 2022-09-28

Foundation item: Supported by the Postgraduate Research and Practice Innovation Program of Nanjing University of Aeronautics and Astronautics (XCXJH20220318)

Biography: HUANG Qiaoyue, female, Master candidate, research direction: pattern recognition. E-mail: huangqy@nuaa.edu.cn

[†]To whom correspondence should be addressed. E-mail: cytang@nuaa.edu.cn

search direction. However, the current recognition techniques for eyes mainly focus on the iris^[4], retina^[5] and other eyeball regions. The data collection of the eyeball area has high requirements for image acquisition, and the subject must be very close to the camera. The subsequent preprocessing and recognition processes are also complicated. Compared with the eyeball area, the processing of the periocular biometrics is relatively simple. It also has a high tolerance for image acquisition and can handle a broad range of distances^[6]. Even though there are masks or other occlusions on the faces, the acquisition and recognition of the periocular images will not be affected. This is extremely suitable for recognition applications during the COVID-19 pandemic. Besides, the periocular features can potentially contribute to significant improvement in terms of distinguishability. For example, it provides the shapes of the eye and eyebrow which contains much biometric information.

In recent years, deep learning has proved to be very effective and popular in computer vision problems. As such, they have been widely explored in face recognition. FaceNet^[7] proposed by Google researchers learns a mapping from face images to a compact Euclidean space where distances directly correspond to a measure of face similarity. Later in 2017, SphereFace^[8] was proposed using the ResNet architecture. Recently, neural architecture search (NAS) has been used in face recognition and achieved outstanding performance^[9]. However, the performance of these models suffers from the covering of facial masks. After the global outbreak of COVID-19, Geng *et al*^[10] introduced a novel Identity Aware Mask Generative Adversarial Network (IAMGAN) to match a masked face with its corresponding full face and achieved an accuracy of 86.5% on Masked Face Segmentation and Recognition (MFSR) dataset. A masked face recognition method^[11] was proposed in 2022. It used Multi-task Cascaded Convolutional Networks (MTCNN) for face extraction and FaceNet for getting the embeddings of the extracted face. The method achieved an accuracy of 94%. In 2021, Huber *et al*^[12] proposed a mask-invariant face recognition solution named MaskInv that aims at producing embeddings of masked faces which are similar to those of non-masked faces of the same identities. MaskInv has enhanced the performance of masked face recognition. However, the aforementioned masked face recognition methods only focus on the deep features extracted by neural networks, which have a heavy reliance on precise and abundant

data.

Thus, in this paper, we propose an algorithm that combines traditional features with deep learning models. On the basis of detecting the masked face and locating the facial landmark points, we segment the periocular region above the mask. After preprocessing, we extract the Local Binary Pattern (LBP), Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradient (HOG) features of the periocular region. The vectors of these three features are used to train the support vector machines (SVM) for face recognition. A deep learning model Angular Visual Geometry Group Network (A-VGG) is proposed to extract deep features and achieve the prediction. Finally, we obtain the decision-level fusion of the four features, which can effectively improve the recognition rate. In daily life, there is also a certain demand for the recognition of the side faces and the images with motion blur. Therefore, in addition to the clear frontal face recognition, we also test the side faces at different angles and the blurred faces. Moreover, the matching between full faces and masked faces is an important task, so we train the faces without masks to recognize the masked faces.

1 Database

1.1 Simulated Masked Face Database

At present, there are few masked face databases, so we add simulated masks to the existing face database. We use the face database published by the Robotics Laboratory of Cheng Kung University, China, which contains 90 subjects, each with 37 images taken from different angles (0° to $\pm 90^\circ$ at 5° intervals). The resolution of images in the dataset is 640×480 . Figure 1 shows the faces at different angles from one subject.

To add masks to face images, we use Dlib^[13] for face detection and face alignment. The face alignment function in Dlib can locate 81 landmark points of a face and number them in order, especially the positions with obvious edge features such as the corners of eyes and mouth. As shown in Fig. 2, the detected face is in the white rectangle and the 81 numbers on the face indicate the locations and the order of the 81 facial landmark points.

After obtaining the coordinates of these facial landmark points, we connect the landmark points around the

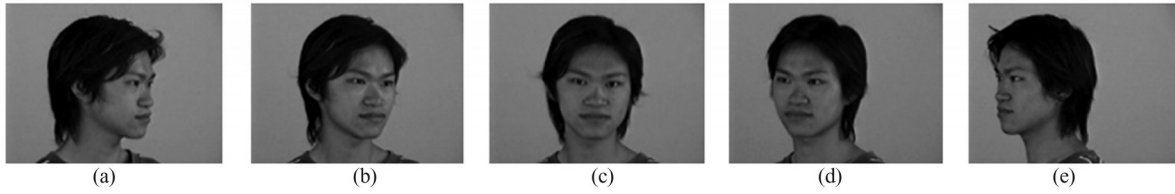


Fig. 1 Faces at different angles in the database

(a) $+60^\circ$; (b) $+30^\circ$; (c) 0° ; (d) -30° ; (e) -60°



Fig. 2 Face detection and face alignment

lower half of the face to define the shape of the simulated mask. Then color filling is carried out inside the outline to obtain the masked face image, as shown in Fig. 3. In this way, a masked face database containing 90 subjects taken from multiple angles is generated.

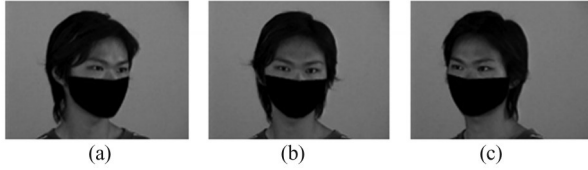


Fig. 3 Masked face at different angles

(a) $+30^\circ$; (b) 0° ; (c) -30°

1.2 Real Masked Face Database

At present, there are few databases of faces wearing real masks. In order to supplement our dataset, we generate our database named HRMF (High-Resolution Masked Faces) by taking masked face images from our friends and schoolmates.

HRMF consists of 70 subjects, each with 4 frontal face images. The resolution of most images is around $3\,000 \times 4\,000$. We captured those images at different times and different locations. Figure 4 shows the sample images of one subject.

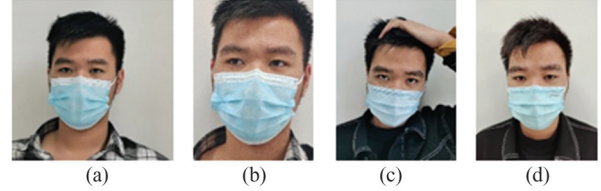


Fig. 4 Real masked face database

(a) (b) were captured for the first time; (c) (d) were captured for the second time

age. It is widely used in face recognition. It can deal with both the simulated and real masked face, as shown in Fig. 5, where the white rectangles show the detected faces. After obtaining the position of the masked face, we use Dlib to complete the face alignment and obtain the coordinates of the facial landmark points.

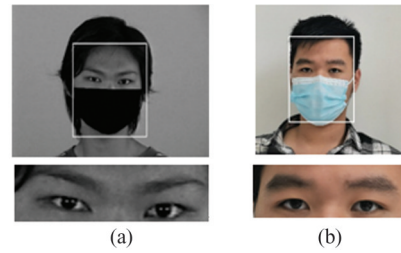


Fig. 5 Segment the periocular region

(a) Simulated masked face; (b) Real masked face

2 Methodology

2.1 Periocular Region Segmentation

In the preprocessing stage, we use the open-source model from the PaddleHub^[14] which is specially trained for masked faces to detect the masked face area in an im-

As shown in Fig. 6, the four points in the red circles are selected to segment the rectangular periocular region, where points 75 and 29 decide the height and points 78 and 79 decide the width. Thus, the periocular image shown in Fig. 5 can be obtained. The segmented periocular area includes eyeballs, eyebrows and the skin around eyes, providing a lot of biometric features for recognition. In the following sections, we will extract and recognize the features of the periocular regions segmented in this way.

2.2 LBP Feature Extraction

LBP^[15] is an operator which describes the local tex-

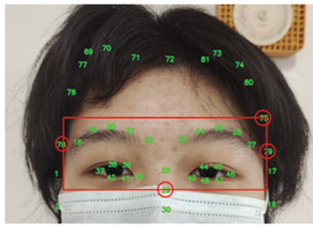


Fig. 6 Facial landmark points

ture features of an image. In a neighborhood of 9 pixels, it compares the gray value of the central pixel with those of other pixels. If the surrounding pixel value is bigger than the central pixel value, the pixel is marked as 1, otherwise, it is 0. By combining these values, a binary number can be generated, which is the LBP value of the cen-

tral pixel and can reflect the texture information around the pixel.

The LBP values form a grayscale image named LBP feature image, with each pixel representing the LBP value of the original image. As shown in Fig. 7, we can find that the LBP operator extracts the texture information of the periocular region. For better recognition results, we use histogram equalization to adjust image intensities and enhance contrast before feature extraction. After image processing and calculating the LBP values, we get the histogram statistics on the LBP feature image and obtain a 1×256 dimensional texture feature vector of the whole image which will be used as the input of the SVM classifier.

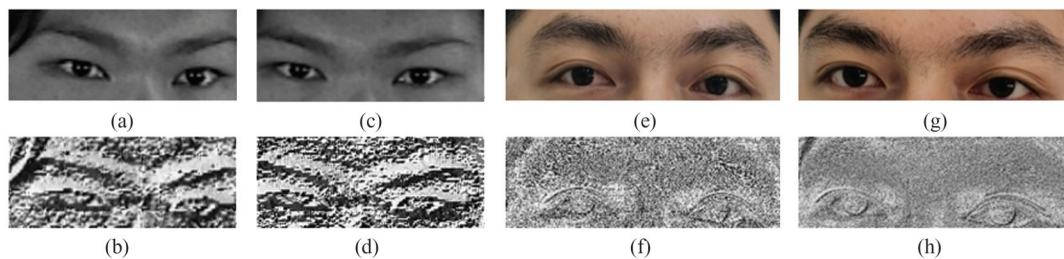


Fig. 7 LBP feature visualization

(a)(c) Original image of subject 1; (b)(d) LBP feature of (a)(c); (e)(f) Original image of subject 2; (g)(h) LBP feature of (e)(f)

2.3 SIFT Feature Extraction

SIFT extracts features based on some key points selected on the object^[16], which is irrelative to the size and rotation of the image. We identify potential key points from the entire periocular region including eyebrows, as shown in Fig. 8, and the small colored circles represent the identified points. The local gradient of the image is calculated in the neighborhood around each key point as the descriptors. A complete SIFT feature vector is generated by connecting all the key point descriptors and its dimension is determined by the number of points. Assuming that the number of identified key points is n , then a SIFT feature vector of $n \times 128$ dimensions can be obtained.

However, because the numbers of key points identified in each image are different, the final feature vector dimensions are different. The vectors with different dimensions cannot be put into the SVM directly. Therefore, we use bag-of-words model^[17] and K-means^[18] for clustering. K-means clustering is carried out on all key point descriptors, thereby k cluster centers are acquired as the visual words which form a visual dictionary. Each key point is mapped to a visual word by finding the nearest center. Then, each image can be represented as a k dimensional vector, where k elements represent the numbers of key points in the corresponding position in the visual dictionary. In this way, we cluster the identified key points of each periocular image and get a new fea-

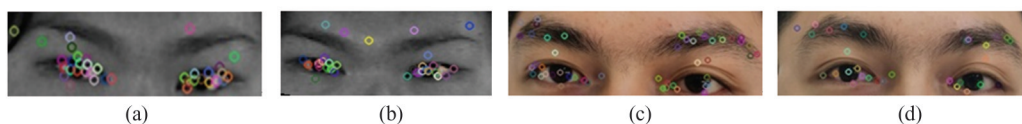


Fig. 8 SIFT key points

(a)(b) subject 1; (c)(d) subject 2

ture vector with a unified dimension. Finally, we put the new vectors into the SVM classifier for training and

classification. The flow chart of the algorithm is shown in Fig.9.

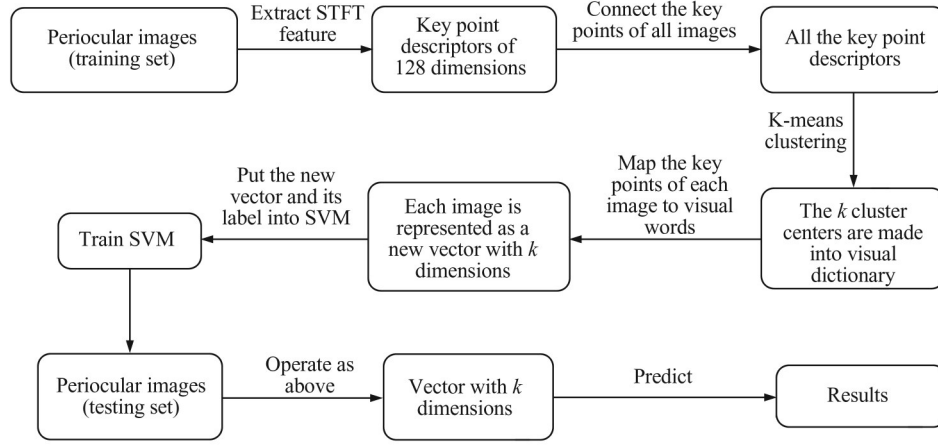


Fig. 9 The flow chart of SIFT+K-means algorithm

2.4 HOG Feature Extraction

HOG (Histogram of Oriented Gradient) forms the feature^[19] by calculating and counting the gradient direction histogram of the local regions of an image. It can maintain good invariance to the geometric and optical deformation of the image. To obtain the feature vector with the same dimension and to improve the recognition rate, the periocular image extracted from the masked face is adjusted to a unified and appropriate size before feature extraction. Then Gamma correction is used to standardize the color space of the input image. After pre-

processing, the size and direction of the gradient are calculated for each pixel.

HOG feature extraction method not only retains the edge information but also retains the directions of edges. We divide the image into $n*n$ cells and group a few cells as a block. The histogram of gradient vectors of each cell is connected and normalized in the block. Then the feature vectors of all blocks are concatenated to get the final HOG descriptor. The dimension of the descriptor is determined by the number of segmented cells and blocks. An example of HOG feature extraction from the periocular image is shown in Fig. 10.

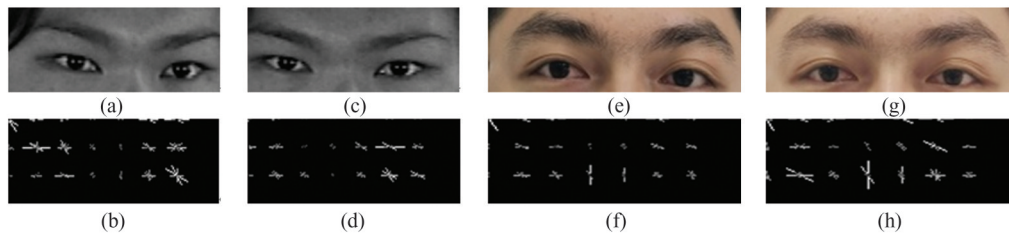


Fig. 10 HOG feature visualization

(a)(c) Original image of subject 1; (b)(d) HOG feature of (a)(c); (e)(g) Original image of subject 2; (f)(h) HOG feature of (e)(g)

2.5 A-VGG Feature Extraction

In recent years, convolutional neural networks (CNNs) have achieved great success in face recognition. It is natural to employ deep learning-based approaches especially CNN for the recognition of masked faces. Since there is not enough labeled image data to train a network from scratch, transfer learning is used in our recognition method. To extract deep features from the in-

formative regions, we have employed a pre-trained model as the feature extractor. VGG16^[20] is a CNN model trained on the ImageNet dataset with the idea of stacked convolution layers of smaller receptive fields. There are 13 convolutional layers, 5 maximum pooling layers, and 3 dense layers which sum up to 21 layers but only 16 weight layers. Its weight configuration is publicly available and has been used in many other applica-

tions.

VGG16 learns face features via Softmax loss. Define the input feature $\mathbf{x}_i$ and its label y_i , and N is the number of training samples. The original Softmax loss can be written as

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i -\log \left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \right) \quad (1)$$

where f is the output of a fully connected layer and in CNN it is just the multiplication of the weight $\mathbf{W}$ and the previous layer output plus bias b . By substituting f , L_i can be reformulated as

$$L_i = -\log \left(\frac{e^{\|\mathbf{W}_{y_i}\| \|\mathbf{x}_i\| \cos(\theta_{y_i}) + b_{y_i}}}{\sum_j e^{\|\mathbf{W}_j\| \|\mathbf{x}_i\| \cos(\theta_{j,i}) + b_j}} \right) \quad (2)$$

in which $\mathbf{x}_i$ and $\mathbf{W}_j$ are the i -th training sample and the j -th column of $\mathbf{W}$, respectively. $\theta_{j,i}$ is the angle between vector $\mathbf{W}_j$ and $\mathbf{x}_i$. However, the original Softmax loss only focuses on separable features. To solve this problem, we use the angular Softmax (A-Softmax) proposed in SphereFace^[8] to enhance the discrimination of features. $\|\mathbf{W}\|$ is normalized into 1, and bias is set to 0. Then the angular margin that can be controlled with pa-

rameter m is incorporated in the loss to learn discriminative features. Therefore, A-Softmax loss can be defined as below.

$$L_{\text{ang}} = \frac{1}{N} \sum_i -\log \left(\frac{e^{\|\mathbf{x}_i\| \cos(m\theta_{y_i})}}{e^{\|\mathbf{x}_i\| \cos(m\theta_{y_i})} + \sum_{j \neq y_i} e^{\|\mathbf{x}_i\| \cos(\theta_{j,i})}} \right) \quad (3)$$

A-Softmax loss has the remarkable effect of high cohesion and low coupling by constraining learned features to be discriminative on a hypersphere manifold. The loss achieves a smaller maximal intra-class distance than the minimal inter-class distance.

After popping out the top output layer, the pre-trained VGG16 can be used to create image embedding vectors. In this way, we transfer the original output layer with Softmax activation to a layer that can extract angular features. By improving the original Softmax loss to A-Softmax, we propose a model, Angular Visual Geometry Group Network (A-VGG), which combines the advantages of VGG16 and SphereFace to learn angularly discriminative features of the periocular region. On the basis of pre-trained convolutional blocks, we fine-tune A-VGG on our dataset to achieve periocular recognition. The model architecture is shown in Fig. 11.

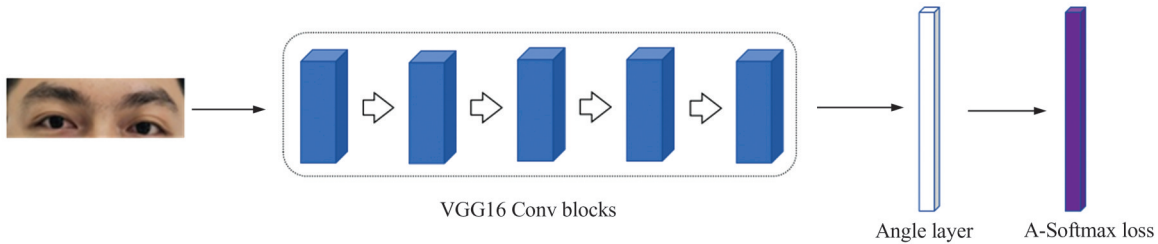


Fig. 11 A-VGG architecture

2.6 Decision Fusion

The decision-level fusion is carried out to obtain the final recognition result. The three traditional feature vectors are extracted and put in SVM for training. SVM is a widely used supervised machine learning model for classification and regression. Basically, SVM finds a hyper-plane that creates a boundary between the types of data. Compared with the newer algorithm like neural networks, SVM has higher speed and it is extremely suitable for a limited number of samples. Thus, we choose it as the classifier of the LBP, SIFT and HOG features. The three trained SVM models are used to predict the labels.

Different from the three traditional features, A-VGG extracts the deep features and computes the labels of test images directly.

Sort the four single feature recognition rates and find the highest one. For each periocular image in the test set, we will obtain four labels predicted by four classifiers. Sort the four labels and find the same labels. The minority is subordinate to the majority to obtain the final recognition result of the masked face image. The label of the classifier with the highest recognition rate will be chosen when the four labels of the test image are different. The process of the algorithm is illustrated in Fig. 12.

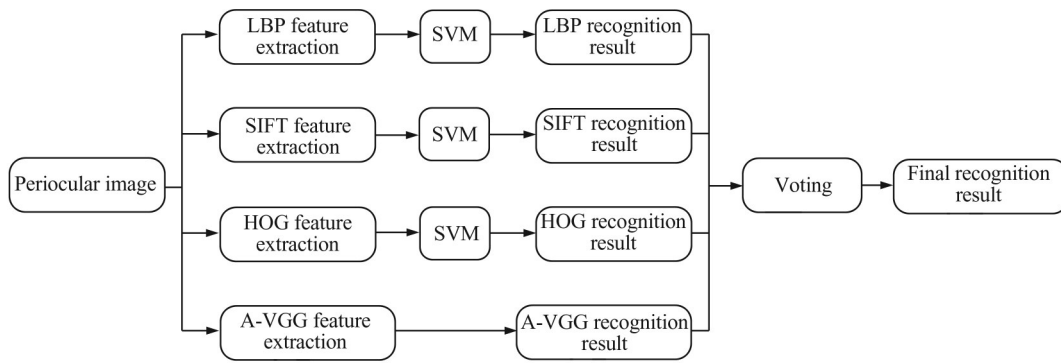


Fig. 12 Decision-level fusion of the proposed algorithm

3 Experiments

Three sets of experiments were carried out to evaluate the proposed algorithm. The first and second were tested on frontal and side simulated masked faces, respectively, and the third was tested on real masked images. The proposed algorithm is compared with some state-of-the-art masked face recognition methods. The first method uses MTCNN and FaceNet^[11] for masked face recognition. The second is a mask-invariant face recognition solution named MaskInv^[12].

3.1 Frontal Face Recognition

In the simulated masked face database, we select the face images of $\pm 10^\circ$ and $\pm 5^\circ$ as the training set and the frontal face images of 0° as the test set. The database consists of 90 subjects, so there are 360 images in the training set and 90 images in the test set. Based on masked face detection and the periocular region segmentation, the feature vectors of LBP, SIFT and HOG are extracted and put into the SVM classifier. We also use A-VGG to extract deep features and output the prediction labels. Each feature is utilized independently for prediction and then, the recognition results of each feature are combined at the decision level.

Besides, to evaluate the robustness of the proposed algorithm, we carry out an experiment on the images processed with motion blur, which simulates the visual streaking or smearing captured on the camera. A processed blurred image is shown in Fig. 13. Since the existing face images usually do not have masks, we also use the original full faces without masks for training to recognize the simulated masked faces. The recognition rates of the four features and the proposed algorithm are given in Table 1. The results of VGG, FaceNet and MaskInv are also given for comparison.



Fig. 13 Fuzzy masked face

Table 1 Frontal face recognition rate

Feature	%		
	Masked face	Blurred masked face	Full face
LBP	91	89	24
SIFT	69	60	52
HOG	80	83	57
A-VGG	99	98	93
VGG	96	94	88
FaceNet	74	62	58
MaskInv	98	96	94
Proposed algorithm	100	99	96

The recognition results show that most feature descriptors have high discrimination and make great progress compared with FaceNet, which means that extracted periocular features can improve the performance of masked face recognition. Among the single feature recognition, A-VGG performs better than all the traditional features and VGG, which means the deep features contain more information than traditional features. Furthermore, it shows that A-Softmax improves the performance of original Softmax in VGG by learning angularly discriminative features. After the decision-level fusion, the recognition rate has been improved to a certain extent compared with single feature recognition and MaskInv. Though the blurred images lead to a small

decrease in recognition rate, the proposed algorithm still maintains its advantage and shows its robustness. When using original full face images to recognize the simulated masked faces, the recognition rate is lower than using masked face images, but the proposed algorithm still has the best performance.

3.2 Side Face Recognition

In our daily life, side face recognition is a highly important task in real-world applications. As for side face recognition, the frontal face images of $\pm 10^\circ$ and $\pm 5^\circ$ of 90 subjects are also selected as the training set, and the face images of -15° , $+20^\circ$, -25° and $+30^\circ$ are selected as the test set. The proposed algorithm is evaluated on the masked faces at different angles. At the same time, VGG, FaceNet and MaskInv trained on masked faces are carried out to identify these side face images for comparison. The experimental result is shown in Table 2.

Table 2 Side face recognition rate

Feature	-15°	$+20^\circ$	-25°	$+30^\circ$
LBP	78	51	43	38
SIFT	64	47	40	17
HOG	67	31	6	7
A-VGG	99	93	86	71
VGG	96	88	83	68
FaceNet	67	61	58	48
MaskInv	93	91	58	52
Proposed algorithm	100	93	86	71

It can be seen from Table 2 that the proposed algorithm also has a higher recognition rate on the masked side faces. However, the periocular regions of the masked side faces at different angles usually have different biometric information. The periocular area images lose more information at large angles. Thus, the recognition rate is lower than that of the frontal faces and decreases with the increase of deflection angle, especially the recognition result of three traditional features. But compared with FaceNet and MaskInv, the proposed decision-level fusion algorithm still has greater advantages. The results show the robustness of our algorithm in side face recognition.

3.3 Real Masked Face Recognition

In HRMF, we put three images of each subject in the training set and one image of each subject in the test

set. The database consists of 70 subjects, so there are 210 images in the training set and 70 images in the test set. Table 3 gives the recognition rates of the single features, the proposed algorithm, FaceNet and MaskInv.

Table 3 Real masked face recognition rate

Feature	Masked face/%
LBP	63
SIFT	41
HOG	79
A-VGG	91
VGG	78
FaceNet	81
MaskInv	86
Proposed algorithm	94

It can be seen from Table 3 that A-VGG also performs best in real masked face recognition. The decision-level fusion improves the recognition rate compared with single feature recognition, including the deep learning model A-VGG. It shows that traditional features have their strength and can be a supplement to deep learning. Though the result of HRMF is not as good as that of the simulated masked face database, the proposed algorithm still makes great progress compared with FaceNet and MaskInv in real masked face recognition.

4 Conclusion

In this paper, we proposed a masked face recognition algorithm. We added simulated masks to a face database and generated a real masked face database, detected and aligned the masked faces, then extracted LBP, SIFT and HOG features to train the SVM classifiers. Then we proposed an improved CNN model A-VGG to achieve periocular recognition. Besides, we tried decision-level fusion based on single feature recognition, which improves the recognition rate to a certain extent. In the four classifiers, A-VGG model is dominant in the final prediction. The final frontal face recognition rate of the simulated masked face reaches 100%. Finally, to evaluate the robustness of the proposed algorithm, we tested it on a side face database and a blurred face database. We also managed to match a masked face with the full face of the same person. Although the recognition rate is lower, it is still better compared with VGG and other existing masked face recognition methods. The research on

masked face recognition is important, especially during the global outbreak of COVID-19. The biological characteristics of the periocular region will have more important research significance.

References

- [1] Kumar M, Dargan S. A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities[J]. *Expert Systems with Applications*, 2019, **143**: 1-27.
- [2] Wang M, Deng W H. Deep face recognition: A survey[J]. *Neurocomputing*, 2021, **429**: 215-244.
- [3] Help Net Security. How well do face recognition algorithms identify people wearing masks? [EB/OL]. [2020-07-28]. <https://www.helpnetsecurity.com/2020/07/28/how-well-do-face-recognition-algorithms-identify-people-wearing-masks/>.
- [4] Zhou W B, Ma X T, Zhang Y. Research on image preprocessing algorithm and deep learning of iris recognition[J]. *Journal of Physics Conference Series*, 2020, **1621**(1): 012008.
- [5] Nagarajan D, Sujatha R, Kavikumar J, *et al.* Retina identification system using machine learning and multiple regression model[J]. *Indian Journal of Public Health Research and Development*, 2019, **10**(7): 178.
- [6] Tan C W, Kumar A. Towards online iris and periocular recognition under relaxed imaging constraints[J]. *IEEE Transactions on Image Processing*, 2013, **22**(10): 3751-3765.
- [7] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering[C]// 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Washington D C: IEEE, 2015: 815-823.
- [8] Liu W Y, Wen Y D, Yu Z D, *et al.* SphereFace: Deep hypersphere embedding for face recognition[C]// 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Washington D C: IEEE, 2017: 6738-6746.
- [9] Zhu N, Yu Z K, Kou C X. A new deep neural architecture search pipeline for face recognition[J]. *IEEE Access*, 2020, **8**: 91303-91310.
- [10] Geng M, Peng P, Huang Y, *et al.* Masked face recognition with generative data augmentation and domain constrained ranking[C]// *MM '20: The 28th ACM International Conference on Multimedia*. New York: ACM, 2020: 2246-2254.
- [11] Sunil T A, Gupta P, Jain A, *et al.* Face recognition with mask using MTCNN and FaceNet[C]// *Artificial Intelligence and Technologies*. Berlin: Springer-Verlag, 2022: 103-109.
- [12] Huber M, Boutros F, Kirchbuchner F, *et al.* Mask-invariant face recognition through template-level knowledge distillation[C]// 2021 *16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*. Washington D C: IEEE, 2021: 1-8.
- [13] Xu M, Chen D, Zhou G. Real-time face recognition based on Dlib[C]// *Innovative Computing: IC 2020*. Berlin: Springer-Verlag, 2020: 1451-1459.
- [14] Github. PaddleHub[DB/OL]. [2021-12-20]. <https://github.com/PaddlePaddle/PaddleHub>.
- [15] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(7): 971-987.
- [16] Park U, Jillela R R, Ross A, *et al.* Periocular biometrics in the visible spectrum[J]. *IEEE Transactions on Information Forensics and Security*, 2011, **6**(1): 96-106.
- [17] Huang T, Ru S R, Zeng Z H, *et al.* Research on motion recognition algorithm based on bag-of-words model[J]. *Microsystem Technologies*, 2021, **27**(4): 1647-1654.
- [18] Bansal M, Kumar M, Kumar M. 2D object recognition: A comparative analysis of SIFT, SURF and ORB feature descriptors[J]. *Multimedia Tools and Applications*, 2021, **80**(12): 18839-18857.
- [19] Dadi H S, Pillutla G K. Improved face recognition rate using HOG features and SVM classifier[J]. *Iosr Journal of Electronics & Communication Engineering*, 2016, **11**(4): 34-44.
- [20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2022-04-10]. <https://arXiv preprint arXiv:1409.1556>.

□