Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Face and occlusion Recognition Algorithm based on Global and Local

To cite this article: Guanwen Yu and Zhaogong Zhang 2020 *J. Phys.: Conf. Ser.* **1453** 012019

Face and occlusion Recognition Algorithm based on Global and Local

Guanwen Yu[1], Zhaogong Zhang[1*]

[1] School of Computer Science and Technology, Heilongjiang University, Heilongjiang Province, China

[*]Corresponding author's e-mail: 2013101@hlju.edu.cn

Abstract: Nowadays, the face recognition direction of unobstructed has been greatly developed, but there is still a huge deficiency in face recognition with occlusion. In order to improve the recognition rate under occlusion, this paper proposes a global and local occlusion recognition algorithm using deep learning for feature learning. Due to occlusion, local significant features will be missing, the impact will cause the accuracy of global recognition to decrease. Therefore, SIFT and SVM algorithms are used to discriminate partial face occlusion, and cut off the occluded local face to form a global face. Global facial features and unmasked local feature are extracted together. In the feature extraction, the LeNet-5 lightweight convolutional neural network and the improved LeNet-5 lightweight convolutional neural network are combined with SoftMax multi-classification. The final classification results use the improved hierarchical voting method to make final decisions on the face, which speeds up the training and prediction speed of the multi-classifier. This experiment will perform occlusion face verification on the AR dataset. The final experimental results show that the method greatly improves the accuracy of occlusion face recognition and has strong robustness to occlusion.

1. Introduction

Face recognition [1-5]is widely used nowadays, has been widely used in video monitoring, identification and other aspects. However, many current face recognition algorithms are applied to face recognition in fixed scenes, only when the recognized face has no occlusion can the accurate face recognition be carried out. However, most face recognition algorithms are affected by physical occlusion, illumination occlusion and posture occlusion. In order to solve the problem of high recognition rate in real scene, it is necessary to solve the problem of reduced accuracy caused by occlusion.

On the issue of occlusion, researchers have never stopped the research on face occlusion recognition, and put forward a lot of algorithms with good effect on occlusion recognition. Wright[6]et al. proposed SRC algorithm in the direction of occlusion recognition, which greatly enhanced the recognition rate of faces with noise and occlusion, SRC algorithm linearly combines all the images in the database to represent the test images. The coefficient vector of the test images is sparse. Then, the sparse representation is solved to judge the categories, and this algorithm can recognize blocked faces well under experimental conditions. However, this algorithm is not accurate in recognizing the changes of facial expressions in real environment, which makes it difficult for SRC algorithm to be applied in real scene face recognition. The above are all traditional blocking recognition algorithms. Yann et al. [7] proposed the deep learning method with lenet-5 as network model for face recognition, Yi Sun et al. 's DeepID[8] algorithm and Yaniv et al.' s DeepFace[9] algorithm adopted the deep learning method for face recognition. However, the accuracy of these methods is higher than that of the traditional algorithm only when the face is not

Journal of Physics: Conference Series **1453** (2020) 012019 doi:10.1088/1742-6596/1453/1/012019

covered, but the recognition rate of the covered face is still not high. Along with the development of deep learning, great progress has also been made in using deep learning to deal with occlusion, Michael Opitz [10] and others defined a new loss function, from the perspective of loss function considering human face detection, and then combined to get the overall detection effect. This paper presents a new CNN depletion layer, called the grid loss, it can be independently minimized convolutional lumps error rate, not the entire characteristic diagram, which leads to local itself more discrimination.

This paper proposes a global and local shadow face recognition algorithm, which first divides the image into four local regions (left eye, right eye, mouth, nose) from global to local, and then uses four local regions in shadow recognition. The algorithm performs shadow recognition on the image, from local to global. To prevent the remaining locations from intersecting the global region and constructing global and local classifiers using these two networks, this paper proposes a weighted voting method based on similar sets to determine the final classification results.

2. Related Work

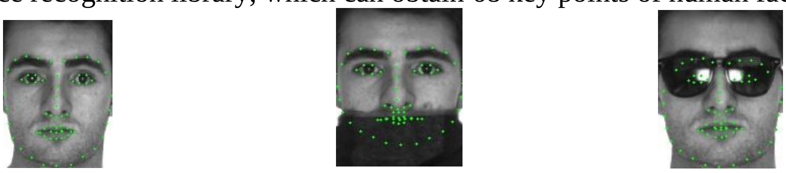
Aiming at the problem of face occlusion recognition, this paper proposes a global and local face occlusion recognition algorithm. The main ideas of this algorithm are as follows:

2.1. The role of global and local features

Global characteristics of the reaction of the whole face of the change of the whole, and the local reaction only the details of the local characteristics of the face, if only to watch the global features, face the different glasses, are likely to lead to face the wrong judgment, if only to watch the local characteristics, the facial expression of the transformation, are likely to face the wrong judgment, so this paper integrated global features and local characteristics of the integrated consideration shade face judgment.

2.2. From global to local

Face_recognition[11] is adopted in this paper to identify key points of human face. Face_recognition is a python-based face recognition library, which can obtain 68 key points of human face.



(a) Unmasked face (b) Scarf covers face (c) Glasses cover the face

Figure 1. Key point detection

In the case of no occlusion, the key points of the face can be accurately detected. In the case of occlusion, the key points of the face without occlusion can be precisely positioned. In the case of occlusion, the key points of the face without occlusion can also be precisely positioned.

Face_recognition is adopted to locate the key points of the face, and four key points (center of left eye, center of right eye, center of nose and center of mouth) are selected, and the key points are cut out with the four key points as the center and the size of 28*28. The four key points are as follows:

A local graph of the cut area is shown in Figure 2 below. The four local graphs will be sent to the subsequent network.

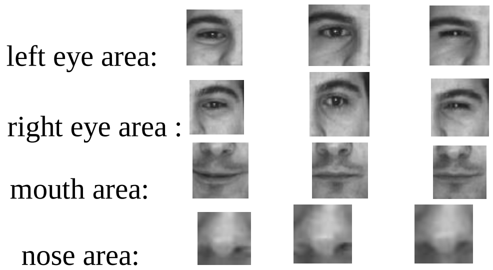


Figure 2. Local face diagram

2.3. Face Occlusion Recognition

Since the discrimination of occlusion mainly detects that the local area is covered or not covered, the problem is a dichotomy problem. Therefore, this paper mainly uses SIFT and SVM to distinguish the four local areas.

2.3.1. The SVM Training. Before SVM training, Will first will have to do all the training images sift algorithm[12] to extract the key points, then set vectors of all key points together, then using the K - Means algorithm combined words of similar visual vocabulary, construct a word list W contains K words, words in the table and then statistics the number of occurrences of each word in the image.

The word list generated above can be used to represent the image, which can be represented as a **K-** dimensional numerical vector. The vector with value of K dimension can be marked and input into SVM for training of occlusion and no occlusion categories.

The whole process of image occlusion discrimination is shown in Figure 3 below:

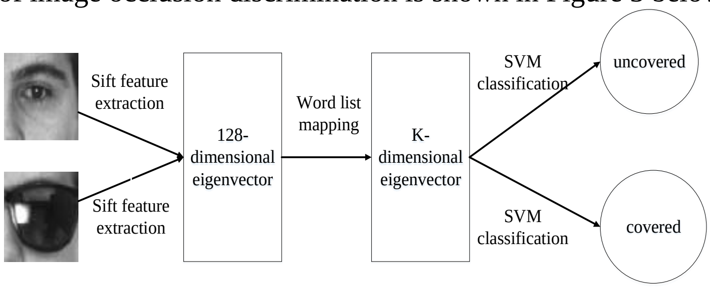


Figure 3.Occlusion discrimination process

2.4. From Local To Global

Will keep out discriminant, covered in directly on the original cross-sectional the block position, because the real scene, shade is generally horizontal block such as wear scarf, wearing sunglasses, so the transverse area of intercepting the ministry branch office, and will give up of this part, leave the noise parts are higher, bigger influence on the results of this part, this part give up after, the rest of the images will be joining together, forming a global facial, concrete process are shown in Figure 4 below.

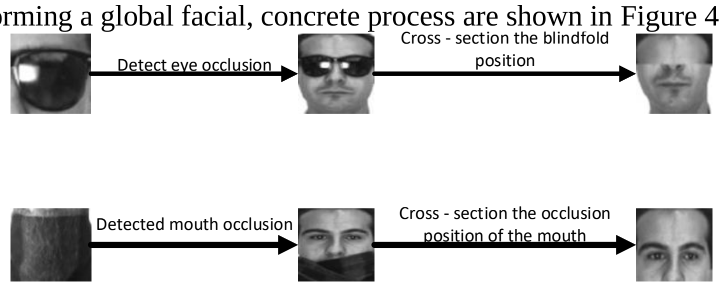


Figure 4.From local to global

Journal of Physics: Conference Series **1453** (2020) 012019 doi:10.1088/1742-6596/1453/1/012019

2.5. Based on global information of deep learning, multiple classifiers are constructed and parallel training is carried out

In this paper, feature extraction will be carried out on local uncovered face blocks and the formed global face, and a classifier will be constructed to classify all blocks. In this paper, a convolutional neural network will be constructed to extract features from all blocks, and SoftMax will be used for multi-classification of features.

In this paper, two network model architectures will be adopted to extract features from images. Lenet-5 network model and improved lenet-5 model will be adopted. Because the pixel value of the input images in this paper is low, lenet-5 model has a low number of layers compared with other models, which can better fit low-pixel images. The model adopted in this paper is shown in Figure 5 and Figure 6.

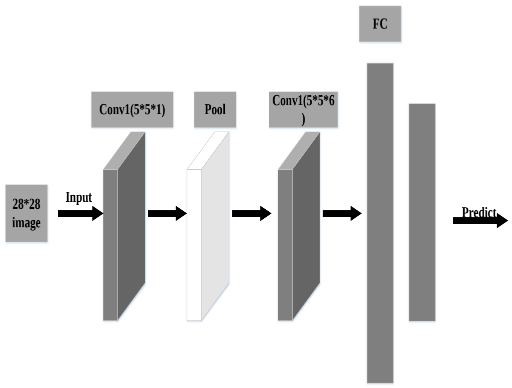


Figure 5. LeNet-5 network structure

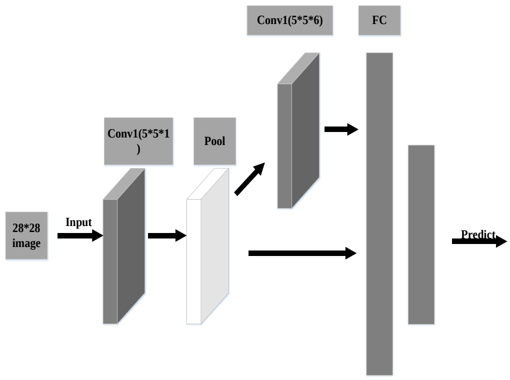


Figure 6.Improved LeNet-5 network structure

2.6. Weighted voting methods on similar sets

This paper proposes a new voting method, weighted voting method on similar sets. When the classifier is trained, each classifier is trained through many data sets, and each classifier will get its own accuracy. This paper will take this accuracy as a reference. The higher the accuracy of the classifier, the more accurate it is, which means that the classification result of the classifier will be more accurate than that of the classifier with low accuracy.

The main process of specific voting strategy is as follows:

(1) In this paper, all results obtained by multiple classifiers are represented by X, and BFPRT algorithm is adopted first. The main idea of BFPRT algorithm is to select the largest or smallest value (k<=n) in the first k from n elements. In this paper, classification results obtained by the first 5 most accurate classifiers can be selected according to the accuracy of the classifier through this algorithm.

$$Y = [X1, X2, X3, X4, X5]$$

X1, X2, X3, X4 and X5 are the results of the five classifiers with the highest accuracy

(2) The results for each category in this article are

$$C^k = [C_k, C_k, C_k, C_k C_k] \quad \text{where } k = 1,2,3 \ldots n$$

n is the number of categories.

The set C composed of the result C^n of each category is:

$$C = \{C^1, C^2, \ldots \ldots, C^n\}$$

Calculate the classification results of Y and the category C used for similarity calculation

$$D(Y, C^k) = \sqrt{\sum_{i=0}^{5} \left(Y_i - C_i^k\right)^2} \quad \text{where } k = 1,2,3 \ldots n \tag{1}$$

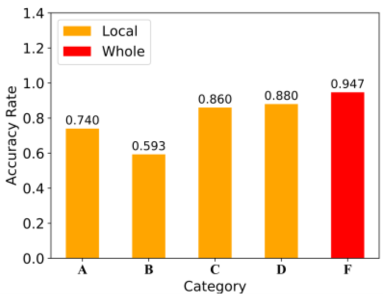
The Euclidean distance is used to calculate the similarity, and the one with the largest similarity will be selected in the end. The class with the smallest Euclidean distance $\mathbf{min}_{(1\le k\le n)}\left(D(Y, C^k)\right)$ is selected as the final category.

3. Experiments

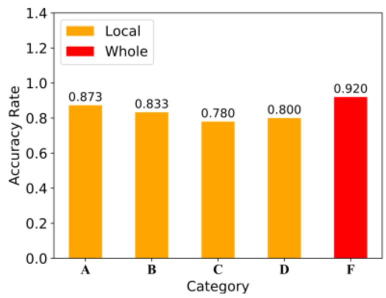
3.1. Occlusion Face Recognition Experiment

In this paper, the data set uses an AR data set with 100 facial images, each person has 26 facial images, including different expressions, different lighting, different occlusions (sunglasses, scarves).The training set will select 7 unobstructed facial images from 50 facial images, these 7 facial images include changes in the expression of the face and changes in the illumination, which can improve the robustness of the face in the real environment under certain conditions. The test set selection will select 3 glasses occluded images and 3 scarf occluded images as a test set. In the selection of the global image of the training set and the test set, the occlusion of the test set is first determined. For the occlusion part, the training set and the test set are cross-cut simultaneously to form a global image, and the unblocked block will form a local block.

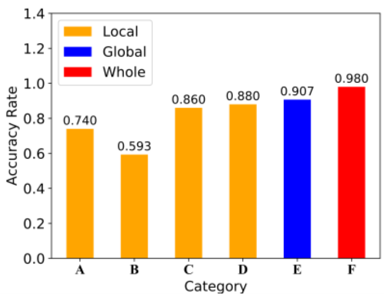
As shown in Figure 7, the glasses occlusion and the scarf occlusion are respectively trained, respectively, the global classifier and the local classifier. This experiment wants to verify the necessity of the global classifier. As shown in Figure 7, the label Local is a yellow strip portion, indicating the performance of the local classifier, and the label Global portion is a blue strip portion, representing the performance of the global classifier, the label part is a red strip, representing the performance of the global classifier and local classifier set in this paper. Shown in (a) and (b) is only the performance of each local classifier, there is no global classifier, and the last red part is only the performance of the local classifier integration.(c) and (d) It can be clearly seen from the figure that after global and local integration, by comparing (a)(b) and (c)(d), it is obvious that the performance of adding the global classifier is compared. There is a significant increase in performance without adding it.



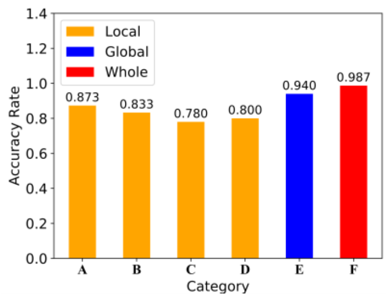
(a) Local classification integration (scarf occlusion)



(b) Local classification integration (glasses occlusion)



(c) Classifier integration (scarf occlusion)



(d) Classifier integration (glasses occlusion)

Figure 7. No global classification and global classifier performance comparison

It can be clearly seen from the figure that there is still a significant gap in performance between the unintegrated global classifier and the integrated global classifier. The performance of the global plus local classifier is significantly higher than the local classification integration.

This article compares the performance of the classifier with dual network and global classifier in the above experiment. Finally choose this strategy as the final strategy of this algorithm, it can be seen from Figure 8 that the accuracy of the algorithm in the occlusion of the scarf is 0.980, and the accuracy of the algorithm in the occlusion of the eye is 0.987, which is compared with other algorithms. The results of the specific experiments are compared as shown in Table 1 .

Table 1. Comparison of the algorithm and each algorithm in this paper.

algorithm	SRC	RSC	RRC	The algorithm in this paper
Sunglasses shade	87．0%	99%	**99.3%**	98.0%
The scarf shade	59.5%	97%	96.5%	**98.7%**

Compared with the above algorithm, the recognition rate of the algorithm in the occlusion of the sunglasses needs to be improved compared with other algorithms, but the recognition rate in the occlusion of the scarf is more accurate than other algorithms.

4. Conclusion

This paper proposes a global and local occlusion recognition algorithm, which mainly uses deep learning to extract features locally, before the feature extraction, the occlusion recognition is needed, and the partial occlusion image is cropped horizontally, and the global image is formed, and the feature is extracted. The soft classification is used for multi-classification. The final result of the classification will adopt an improved voting method. The above is the main content of the algorithm. The detection results of the algorithm show that there is a high accuracy in the occlusion of the real environment. However, it is necessary to further study how local and global integration is better for face recognition. Increasing the speed of recognition will be a challenging task.

References:

[1] Peng, C. , Gao, X. , Wang, N. , & Li, J. . (2015). Graphical representation for heterogeneous face recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence, 39(2), 301-312.
[2] Cao, Q., Li, S., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). VGGFace2: A Dataset for Recognising Faces across Pose and Age. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (pp.67-74).
[3] Jin, Y. , Lu, J. , & Ruan, Q. . (2017). Coupled discriminative feature learning for heterogeneous face recognition. IEEE Transactions on Information Forensics and Security, 10(3), 640-652.
[4] Duan, Y. , Lu, J. , Feng, J. , & Zhou, J. . (2017). Context-aware local binary feature learning for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, PP(99), 1-1.
[5] Ye, Y. , Chen, L. , Hou, S. , Hardy, W. , & Li, X. . (2018). Deepam: a heterogeneous deep learning framework for intelligent malware detection. Knowledge and Information Systems, 54(2), 265-285.
[6] Wright J,Yang A Y,Ganesh A,et al. (2008). Robust face recognition via sparse representation[J].IEEE Transactions on Pattern Analysis and Machine Intellingence, 1(2):210-227.
[7] Lecun, Y. , Bottou, L. , Bengio, Y. , & Haffner, P. . (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
[8] Yi, S., Wang, X., & Tang, X. (2014). Deep Learning Face Representation from Predicting 10,000 Classes. IEEE Conference on Computer Vision & Pattern Recognition.
[9] Taigman, Y. , Yang, M. , Ranzato, M. , & Wolf, L. . (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society.
[10] Opitz, M. , Waltner, G. , Poier, G. , Possegger, H. , & Bischof, H. . (2016). Grid loss: detecting occluded faces.
[11] Kazemi, V. , & Sullivan, J. . (2014). One Millisecond Face Alignment with an Ensemble of Regression Trees. 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE.
[12] Lowe, D. G. . (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91-110.