International Journal of Computational Intelligence Systems and Applications. Volume 9, Issue 1, (2022) 37-44
http://www.cayley-nielson.com/ijcisa/

An Algorithm based on DICW for Face Recognition under Occlusion

Lin Shi[1], Zhigang Li[2]
[1] School of Artificial Intelligence, North China University of Science and Technology, Tangshan, China;
[2]Computer Center, TangShan College, TangShan, China
[1]Corresponding author
E-mail: [1]*12389248@qq.com*, [2]*47971396@qq.com*

Abstract

Based on the robust sparse coding method (RSC)and the dynamic image-to-class warping method (DICW), it proposes an improved algorithm called RSD, using RSC to get a similar set of human face images, and then DICW is used for accurate recognition to get the final recognition result. It is concluded that when the number of face image samples exceeds a certain value, the face image in the training set is occluded, the face image in the test set is not occluded(O-U) , and the face images in the training set are not occluded, the face images in the test set are occluded(U-O), and the face images in the training set and the test set are occluded(O-O), the recognition rate of the improved RSD algorithm is 3% ~ 5% higher than that of the other two methods in separate experiments, which verifies the effectiveness of the improved RSD algorithm.

Keywords: Face recognition under occlusion, RSC, DICW

1. Introduction

With the rapid development of automatic face recognition (AFR) technology, the application of face recognition has become more and more extensive, going deep into transportation, communication, national defense, access control and verification information [1]. However, because the face image is complex and often occluded by different degrees (such as masks, scarves, glasses, etc.), it greatly increases the difficulty of face recognition. The problem of face recognition under blocked conditions is an urgent problem to be solved, and it is the most critical step that hinders face recognition from moving towards a wider application field [2].

S. M Yoon et al. [3] proposed the definition of occluded face for the first time. Occluded face refers to the face that is occluded by sunglasses, scarves, masks, ornaments, hair and other objects, resulting in the inability to accurately recognize or classify faces.

Traditional face recognition methods mainly include methods based on template matching, methods based on geometric features, methods based on artificial neural networks and face recognition methods based on sparse representation [4].

Methods based on heuristic discrimination rules usually solve specific occlusions, mainly using pre-determined rules to detect these specific occlusions [5,6], and then face recognition under occlusion. Lutz Goldmann et al. [7,8] combined the knowledge of statistics and machine learning and proposed a recognition method based on face organs. However, the current algorithm generally has low recognition rate and poor stability [9-11].

Therefore, this paper proposes an improved RSD algorithm to improve the recognition rate and stability of face under occlusion.

2. Methods

2.1. Framework of the model

The purpose of the improved RSD algorithm is to improve a fast matching face recognition method under occluded conditions in combination with the sparse representation method to solve the problem of slow running speed of local image processing algorithms. This method first determines a similar target set quickly by using the robust sparse representation classification method, and then uses the dynamic image-to-class warping method(DICW) to accurately recognize the similar target set, this not only improves the recognition accuracy of sparse representation method in some cases, but also improves the recognition speed of DICW method, thus improving the recognition rate and speed of occluded face recognition method. As shown in Fig.1, it is the model of improved RSD algorithm.

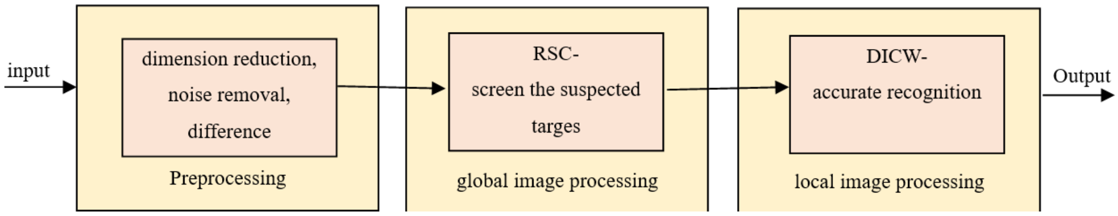

Fig.1 Framework of the model of improved RSD algorithm

2.2. Implementation process

The implementation process of the improved RSD algorithm is mainly divided into three steps: the first step is the preprocessing of face images, which aims to eliminate redundant information, reduce the amount of calculation and accelerate the experimental speed of the algorithm. The second step is to use the advantages of RSC algorithm to identify quickly and get a similar set. In the third step, the DICW method is used to identify the similar sets accurately, and finally the experimental results are obtained. Construct a training data set from the face image database, assuming that there are I different people, and the i_{th} person has n_i samples. All samples are dimensionally reduced first, and then vectorized to generate an m-dimensional column vector. Finally, these column vectors are arranged in rows to form a training dictionary D. as shown in Fig. 2, the construction flow of the training dictionary D can be expressed by formula(1):

$$D = \left[D_1, D_2, \dots \dots D_I, \;\; D \in R^{m \times n}\right] \quad (1)$$

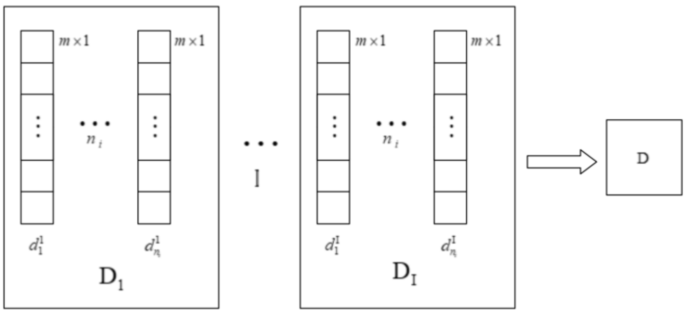

Fig.2 Training dictionary

Where,$Di = [d_1^i, d_2^i, \dots d_{n_i}^i]$,$Di \in R^{m \times n_i}$, $d_{n_i}^i$refers to the vector after the dimension reduction of the n_i face image of the i_{th} person. $n = n_1 + n_2 + \cdots + n_I$.use the training dictionary D to solve the solution of the measured target y, that is: $y = Da$, $a \in R^m$, generally, the equation $y = Da$ is indeterminate and has innumerable solutions.

When it's explained from the essence of face recognition, the solution is sparse and there are non-0 values only in a few places, and these non-0 values represent the important feature information of the recognized face, then face recognition is transformed into finding the sparse solution of $y = Da$,$a \in R^m$. Because there is noise, plus the noise vector, then change the equation to: $y = Da + z$,$z \in R^m$, $\|z\|_2 < \varepsilon$ denotes a noise unit with limited energy, then the problem becomes:

$$\hat{a} = arg\min_a \|a\|_0 \, s.t. \|Da - y\|_2 \le \varepsilon \tag{2}$$

Using the theory of compressed sensing, the problem can be transformed into:

$$\hat{a} = arg\min_a \|a\|_1 \, s.t. \|Da - y\|_2 \le \varepsilon \tag{3}$$

There are many ready-made methods to solve the above optimization problems, in order to improve the performance of the algorithm, Homotopy is used to find $\hat{a}$, and the coefficient $w_i(\hat{a})$, which is only related to the i_{th} person in $\hat{a}$ is extracted. Other positions are zero, and each individual reconstruction error is:

$$R(y) = \{r_i(y) | r_i(y) = \|y - Dw_i(\hat{a})\|_2^2, i = 1,2, \dots I\} \tag{4}$$

Select j persons with the smallest reconstruction error as the similar set, and according to the reconstruction error，mark the sequence from small to large, and record it as $s_1, s_2, \dots s_J$, then form the suspected target set $S = \{s_1, s_2, \dots s_J, \}$, if$\left|r_{s_1}(y) - r_{s_2}(y)\right| > \tau$, τ is the preset threshold, then it can be directly determined that the target sample is the person represented by s_1.Otherwise, enter the stage of accurate identification.

The DICW method first blocks the face image, and divides the object y into B blocks according to certain rules:$y = \{y_1, y_2, \dots y_B\}$, where y_b represents the pixel value vector of the b_{th} block image of the recognized face image. Then, all the face images in the suspected target $s_j, j = 1,2, \dots J$are divided into C blocks:$d_k^{s_j} = \{d_{k,1}^{s_j}, d_{k,2}^{s_j}, \dots d_{k,c}^{s_j}\}$, $d_k^{s_j}$ represents the pixel image vector of the c_{th} image of the $k_{th}, k = 1,2, \dots, n_{s_j}$ face of the s_{jth} individual. Then extract the first-order difference features: if a face is divided into F blocks, $g(x, y)$ represents the pixel value at the coordinates (x, y) in the image, $g_f(x, y)$ represents the pixel value matrix of the f_{th} block. The difference between $g_f(x, y)$ and its directly adjacent block $g_{f+1}(x, y)$is used to approximate the first-order differential features, that is:$\Delta g_f(x, y) = g_{f+1}(x, y) - g_f(x, y)$. Then, a new first-order difference eigenvector is used to replace the initial image feature vector, forming $D_S = [D_{S_1}, D_{S_2}, \dots D_{S_j}]$, when DICW algorithm is used for matching, each time only compare and identify with the data in a similar target set. Find a regular path that can minimize the cost function. The regular path w represents the match of the measured target y and the suspected target D_{S_j} on the time axis, where $W = \{w(1), w(2) \dots w(T)\}$, $w(t) = Match(b; c, k)_t$which indicates that y_b and $d_{k,c}^{s_j}$ is matching at time t, where $b = 1,2, \dots B$, $c = 1,2, \dots,$ C, $\max(B, C) \le T \le B + C - 1$. For DICW algorithm, it's required that b, c should meet certain constraints to restrict the forward direction and the increased step size, that is, when $t = 1, b = c = 1$,when $t = T, b = B, c = C$. When $t \ne 1, T$, if it's (b, c)at t, it's one of the $\{(b-1, c), (b, c-1), (b-1, c-1)\}$ at $t - 1$.

$$E(W) = \sum_{t=1}^{T} V_{w(t)} \tag{5}$$

The cost function is as shown in formula (38), $V_{w(t)}$ is the Euclidean distance between y_b and $d_{k,c}^{s_j}$ at time t according to $w(t)$ .When the paths are different, the cost will be different, so the optimization purpose of the algorithm is to solve the optimal regularization path:

$$\widehat{W} = arg \min_{w} V_{w(1)} + V_{w(2)} + \dots + V_{w(T)} \tag{6}$$

Make the cost function get the minimum value $d_{dicw}^{s_j}\left(y, D_{s_j}\right) = E(\widehat{W})$, $d_{dicw}^{s_j}\left(y, D_{s_j}\right)$ means the similarity measure between measured target y and suspected target s_j. As shown in Tab. 5, it is the flow chart of the improved RSD algorithm.

Tab.5 Flow chart of the improved RSD algorithm

Algorithm flow
Input: face image data preprocessing, dimension reduction, construction of training dictionary $D = [D_1, D_2, \dots \dots D_I,\ \ D \in R^{m\times n}]$
Step 1: use the D to find the sparse solution of y : $\hat{a} = arg \min_{a} \|a\|_1\ s.t.\ \|Da - y\|_2 \le \varepsilon$
Step 2: use sparse solution a to solve the reconstruction error: $R(y) = \{r_i(y) \mid r_i(y) = \|y - Dw_i(\hat{a})\|_2^2, i = 1,2, \dots I\}$
Step 3: construct suspected target set: $S(y) = \{s_1, s_2, \dots s_J,\}$ if$\left\|r_{s_1}(y) - r_{s_2}(y)\right\| > \tau,$ directly determine, otherwise enter DICW module
Step 4: select the classes in the similar target set, segment and extract the first-order difference features: $d_k^{s_j} = \{d_{k,1}^{s_j}, d_{k,2}^{s_j}, \dots d_{k,c}^{s_j}\}$ $\quad \Delta g_f(x,y) = g_{f+1}(x,y) - g_f(x,y)$
Step 5: calculate the minimum value of the cost function using the dynamic image-to-class warping algorithm: $G_{b,c,k} = V_{W(t)} + \min\{G_{b-1,c-1,:}, G_{b-1,c,:}, G_{b,c-1,:}\}$ $d_{dicw}^{S_j}\left(y, D_{s_j}\right) = \min\{G_{B,C,1:K}\}$
Output: $identy(y) = \arg\min\{d_{dicw}^{s_1}, d_{dicw}^{s_2}, \dots d_{dicw}^{s_J}\}$

Algorithm execution: using dynamic image-to-class warping method, initialize a three-dimensional accumulation matrix$G \in R^{B+1\times C+1\times K}$, suppose that $G_{0,0,:} = 0, G_{1:B,1:C,:} = +\infty$,where$G_{b,c,k}$ at time t is equal to $V_{w(t)}$ plus the smallest one of all accumulated values at time $t-1$,that is:

$$G_{b,c,k} = V_{W(t)} + \min\{G_{b-1,c-1,:}, G_{b-1,c,:}, G_{b,c-1,:}\} \tag{7}$$

Calculate from $t = 1$to $t = T$ in sequence, and then get the similarity measure of the measured target y and the suspected target s_j as follows:

$$d_{dicw}^{S_j}\left(y, D_{s_j}\right) = \min\{G_{B,C,1:K}\} \tag{8}$$

Calculate the similarity between the measured object y and the images in the similar set, then the identity of Y is:

$$identy(y) = \arg\min\{d_{dicw}^{s_1}, d_{dicw}^{s_2}, \dots d_{dicw}^{s_J}\} \tag{9}$$

3. Results

3.1. Analysis of experimental results

In this experiment, the improved RSD algorithm is tested on the AR face image database.

The experimental environment is 32-bit Win10 system, and the language is Python 3.6. The integrated development environment is pycharm. Some frameworks (Tensorflow, Opencv) and packages (Tensorboard, Numpy, Matplotlib and Dlib) are used in the experiment.

The experiment consists of four parts, U-U, U-O, O-U, and O-O. since the AR face image database is composed of two sets, session1 and session2, which were captured in different time periods and include changing factors such as lighting, expression and occlusion in natural state, so they can reflect live face samples in real life.

In the four parts of this experiment, each person has k face sample images, and k is from 1 to 7, the performance of each algorithm will be compared during the experiment, and homotopy algorithm will be used in the experiment.

Set the regularization parameter $\lambda = 1e-5$, $\tau = 0.5$, set $B = C$ and the size of the block is 5×5. The experimental results are as shown in Fig.3-Fig.6:

In Fig.3, it can be found that the recognition rate of the algorithm will increase with the increase of the number of samples for each person. When the number of samples is 1 to 4, the recognition rate of SRC method is slightly higher, but when the number of samples continues to increase to 5, the recognition rate of SRD method has made a breakthrough, surpassing other algorithms.

It can be found from Fig.4 (-1represents glasses and -2 represents scarf) that for the occlusion caused by glasses and scarves, the influence of scarves occlusion is greater than glasses, that's because the occlusion area caused by the scarf is large in general, so the algorithm is affected. It can also be found that the recognition rate of RSD algorithm is not as high as that of DICW algorithm for the glasses occlusion, but the RSD algorithm has a better average recognition rate than the DICW algorithm in these two kinds of occlusion, which also shows the stability of RSD algorithm. And with the increase of the number of samples, the recognition rate of the algorithm will be promoted.

It can be seen from Fig. 5 that the recognition rate of RSD algorithm is similar to that of DICW algorithm, both slightly higher than the recognition rate of SRC algorithm. In this O-U case (shown in Fig.5), the recognition rate of the algorithm is generally lower than in the U-O case (shown in Figure 4). This is because if there is occlusion in the face image of the training set in the process of feature extraction, it is impossible to accurately extract the features of the face, and it is even more impossible to recognize the face, so the recognition rate will decrease. But with the increase of the number of samples, the recognition rate of the algorithm has improved.

It can be seen from Fig. 6 that the recognition rate of RSD algorithm is similar to that of DICW algorithm, and the recognition rate of both algorithms is higher than that of SRC algorithm. With the increase of the number of samples, the recognition rate of the algorithm has improved. And the recognition rate of the algorithm tends to be stable when the number of samples per person exceeds 3.

Through the comprehensive comparison of the four experiments and the analysis of the experimental data, it is concluded that the recognition rate of the algorithm will be higher with the increase of the samples. The recognition rate of RSD algorithm is higher than that of other algorithms to varying degrees.

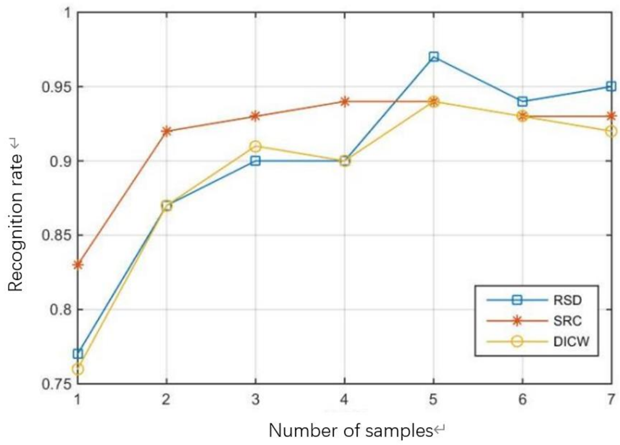

Fig.3 Recognition rate of different methods in U-U scene

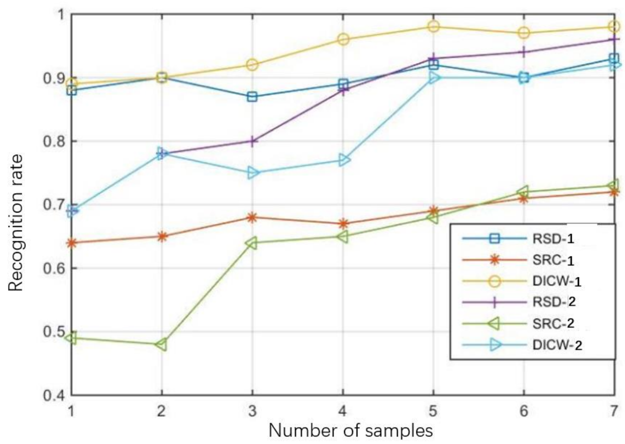

Fig.4 Recognition rate of different methods in U-O scene

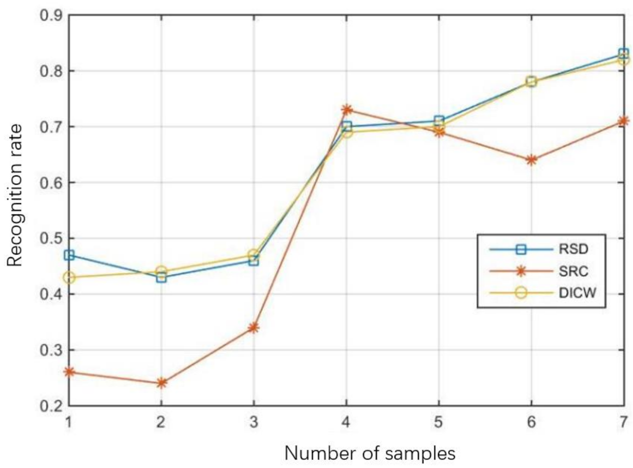

Fig.5 Recognition rate of different methods in O-U scene

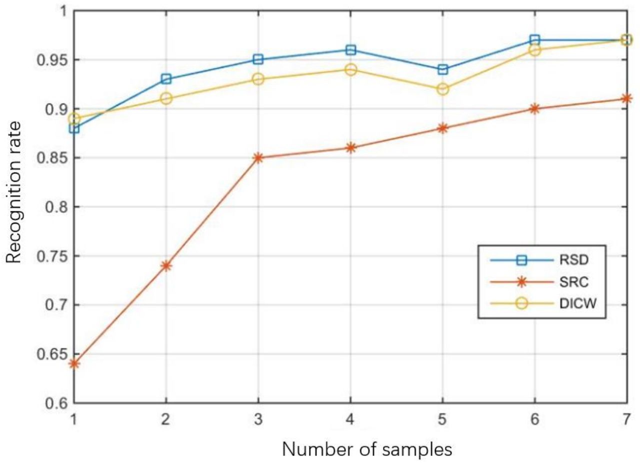

Fig.5 Recognition rate of different methods in O-O scene

4. Conclusion

We propose an improved RSD algorithm, firstly, the robust sparse coding method is used for rough recognition to get a similar set of human face images, and then the dynamic image-to-class warping method is used for accurate recognition to get the final recognition result.

The experiment verifies the improved RSD algorithm and analyzes the data obtained from the experiment. The results show that the recognition rate of the RSD algorithm has been significantly improved when the face images in the training set are not occluded and the face images in the test set are occluded(U-O), and the recognition speed has also been significantly improved compared with DICW method alone.

The proposed improved RSD algorithm has high application value. Restricted by time and technology, the occluded face recognition method in this paper still needs improvement. In the subsequent research, the occluded area in the face image can be increased, and then the algorithm can be improved to improve the robustness of the algorithm.

References

[1] X Lu. "Research and implementation of occluded face recognition algorithm based on sparse representation". *North China Electric Power University*, (2016):23-26.

[2] W. J. Shen. "Research on feature extraction and recognition of occluded face images". *Zhejiang University of Technology*, (2015):20-24.

[3] J Lu, Y.P Tan, G Wang. "Discriminative Multimanifold Analysis for Face Recognition from a Single Training Sample per Person". *IEEE Transactions on Pattern Analysis & Machine Intelligence,* 35 (1) (2013):39-51

[4] CH Qian. "Research on face recognition technology based on deep learning". *Southwest Jiaotong University,* (2017):25-34

[5] Y.S Guo. "Research on occluded face detection algorithm for ATM", *Nanjing University of Technology*, (2013):30-36.

[6] H. Zhang. "Homologous video retrieval and trademark identification", *Beijing University of Posts and Telecommunications*, (2011):38-42.

[7] J Li, W Hao, X Zhang. "Learning Kernel Subspace for Face Recognition". *Neurocomputing,* 151(P3) (2015):1187-1197.

[8] G Guo, S Z Li, K Chan. “Face Recognition by Support Vector Machines”. *IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, (2000): 196-201.

[9] F. Liu. “Research on face recognition algorithm based on sparse representation”. *China University of mining and technology*, (2017):56-60.

[10] X.F Fu, Y Zhang, J Wu. “Joint aided dictionary learning and low rank decomposition face recognition under occlusion expression change”. *Chinese Journal of image and graphics,* 23(03), (2018):399-409.

[11] K Zhang, Z Zhang, Z Li. “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks”. *IEEE Signal Processing Letters*, 23(10),2016:1499-1503.