A Novel Sparse Representation Classification Face Recognition Based on Deep Learning

Junying Zeng, Yikui Zhai, Junying Gan
School of Information Engineering, Wuyi University, Jiangmen, China
zengjunying@126.com

Abstract—**The existing face recognition under pose and illumination variations is a challenging problem. A novel sparse recognition face recognition algorithm based on deep learning is presented in this paper. The deep learning network extracted global and local information, the deep learning network adopted the supervised Convolution restricted Boltzmann machine. The features extracted could recover the face image and reduce intra-identity variances, while maintaining discriminativeness between identities. The algorithm obtained the feature by the deep network and realized fast sparse classification by smoothed l_0 norm. Experimental results on FERET face database show that the proposed algorithm can improve recognition rate and recognition speed when dealing with various conditions such as pose variation.**

Keywords- ***face recognition; deep learning; feature extraction; sparse classification; l_0 norm***

I. INTRODUCTION

Face recognition (FR) is an active research topics in biometric identification, has a wide range of applications, such as access control, video surveillance, etc. But in actual applications, face recognition is affected by many factors, for example, face images were collected in the monitoring system, low resolution, pose variation, complex illumination, which result in very low human face recognition rate. Pose, illumination variation in face recognition is introduced into the nonlinear parameters, but the existing pattern recognition methods, such as the shallow layer structure of the support vector machine (SVM), the representation of the complex functions, complex classification problems have obvious deficiencies. But deep learning can extract the essential feature of the input data through nonlinear network learning, and realize the representation of a complex function [1-2].

The sparse representation theory proposed (SRC) by John Wright Allen y. [3] adopt the overall feature extraction methods, such as Eigenface and Fisherface [4-6]. When the samples are enough, SRC can achieve good recognition effect when dealing with the problem such as illumination, pose variation. In reality, enough training samples were difficult to obtain, the extracted features are affected by illumination, pose, and the linear relationship between the sample are destroyed, which lead to low recognition rate. Recently, in the literature [7] by zhang proposed collaborative recognition classification (CRC), which use the l2 norm standard coding coefficient instead of the time-consuming l1 norm, can achieve the same recognition rate. But at the same time, many literature [8-9] use smoothed l_0 algorithm to realize the SRC, smoothed l_0 algorithm approximate by a smooth continuous function. Smoothed l_0 algorithm has the more fast calculation speed and requires fewer measured values than the other sparse representation method. In view of the insufficient of the existing face recognition methods under pose variation, a novel algorithm based on the deep learning network is proposed in this paper. The algorithm extract face feature by deep network and classify by using l_0 norm. The algorithm extract deep feature from arbitrary pose and illumination of face images. The new method reduce facial differences between the same class face images, while maintaining the difference between different class so as to eliminate the differences under pose variation., as shown in figure 1. In addition, the LBP, Gabor feature[10-12] can not restore the original images, but the extracted features based on deep learning can reconstruct the frontal face of the same identity person under standard illumination. The experimental results on FERET face database show that the algorithm still can maintain a higher human face recognition rate and fast recognition time when big pose variation.

The rest of the paper is organized as follows: section 2 briefly introduces the deep network feature extraction face image algorithm and SRC algorithm. A sparse classification recognition algorithm based on deep learning network and l0 norm is present in Section 3. Section 4 provide the experimental results on the FERET face database and the effectiveness of the algorithm is analyzed and verified under pose variation, followed by conclusions in section 5.

II. RELATED WORK

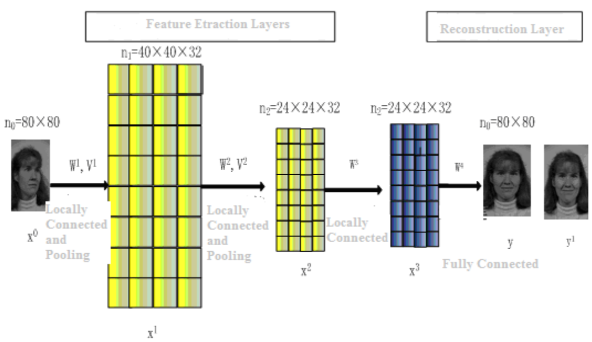
In 2006, the machine recognition scholar Geoffrey Hintonn published a paper [1] in the journal science and proposed deep learning theory. The paper proposed the idea of deep learning based on the deep belief network(DBN) and solve the deep neural network optimization problem. Deep learning has conforms to the human brain structure hierarchical organization ideas and concepts, in a big data environment recognition speed, has the advantages of good recognition effect.

Human accept outside signal through retinal. Through neural network, the retina will send the see figure to the brain through layer upon layer of transmission way. At the same time, in the process of signal transmission, the brain will extract the different feature information and constitute the features of different level. Deep model is similar to the human brain, by the stack many hidden layer, and the hidden layer is intelligent trained in unsupervised way. For example, the deep belief networks and deep Boltzmann machine(DBM) stack many

978-1-4673-7211-4/15 $31.00 © 2015 IEEE
DOI 10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.345

layers of restricted Boltzmann machines(RBM) and can extract different levels of features. In recent years, the convolutional restricted Boltzmann machine (CRBM)[13] was proposed, which incorporates local filters into RBM. The learned filters can preserve the local structures of data. Unlike DBN and DBM employ fully connected layers, the proposed deep network combines both locally and fully connected layers, which enables to extract both the local and global information. And the proposed deep network with a supervised scheme and the extracted features can recover the frontal face image.

Figure 1 is a deep network structure of the model. The input image is arbitrary pose and illumination of face image, the feature extraction of the deep network layer includes three connection and two pool layer. First, the input images were coded by feature extraction layer, there is three local connection layer and two pool layer and alternately stacked. Each layer extract face feature in different scales. For example, as shown in figure 1, the first layer outputs 32 feature mapping. Every images outside the face region has a lot of strong response, which the main information of face images capture face pose, illumination, and some of the strong response in the area of the face, are use to capture facel structure. In the second local connection layer which output feature mapping, a strong response in the face region has been significantly reduced, which suggest that discard most face differences and keep the face structures. The third output features local connection layer, the feature is sparse and keep identity. And using the complete connection layer, the extrated features can restore the standard view of face image.



Figrue.1 Deep learning network (include feature extraction layer and reconstruction layer)

The application of sparse representation in face recognition is proposed in reference [1]. Suppose there are C subjects and M samples from each subject, the training vectors for each subject are denoted by $A_k = [v_{k,1}, \cdots, v_{k,j}, \cdots v_{k,M}]$, $k = 1, \cdots, C$ and $j = 1, \cdots, M$. Forming a dictionary of training vectors from all available samples as $A = [A_1, \cdots, A_k, \cdots, A_C] = [v_{1,1}, \cdots, v_{k,j}, \cdots, v_{C,M}]$, and linear representing testing sample y as

$$y = Ax_0 \tag{1}$$

where $x_0 = [0, \cdots, 0, x_{k,1}, \cdots, x_{k,M}, 0, \cdots, 0]^{\mathrm{T}} \in \Re^{MC}$ is a ideal solution coefficient vector whose entries are zero except those associated with the testing sample. Since the solution for-med by using a small number of training vectors from the large training samples, the solution is sparse and can be obtained by

$$\min \| x \|_1 \quad s.t. \quad y = Ax \tag{2}$$

In reality, due to the influence of various factors, the solution obtained from Eq. (2) will notconsist solely of vectors from a single subject. At this moment, the subject of testing sample can be identified by

$$y \in \arg\min_i \gamma_i(y) = \| y - A\delta_i(x) \|_2, i = 1, 2, \cdots, K \tag{3}$$

where $\delta_i(x)$ is a characteristic function. It indicates that the rest of coefficients in x are set to zero except for the kth . Occlusion, erosion, disguise and so on of part of a face image make the classification problem more difficult and the framework outlined above can be modified. For example

$$y = Ax_0 + \varepsilon \tag{4}$$

where ε represents disguise or occlusion.

Smoothed l_0 algorithm model is $\min \|\mathbf{x}\|_0 \ s.t. \ \mathbf{y} = \mathbf{A}\mathbf{x}$, in order to solve the l_0 norm minimum, gradient projection method and the steepest descent method is used. Continuous gaussian function is used to approximate the vector l_0 norm, the vector of l_0 norm can be considered as a discrete function, therefore, the solved sparse solution of the continuous gaussian function is made to minimize the norm of this vector. Smoothed l_0 algorithm selects a standard gaussian function, the function can be seen as a smoothed continuous function, it can make l_0 norm minimize through the function approximation, wherein, the standard gaussian function $f_\sigma(x_i) = e^{-\frac{x_i^2}{2\sigma^2}}$, x_i represent a certain weight vector X , and σ as a parameter, defined as

$$F_\sigma(x) = \sum_{i=1}^{N} f_\sigma(x_i) \tag{5}$$

when σ value is minimum and maximizes the F_σ value, minimum l_0 norm value can be obtained. The steepest descent method is that Formula $x \leftarrow x + \lambda_k \nabla F_a(x)$ through iterative , the reduced σ resulting in reduced λ_k , and the smaller the value σ , the greater the fluctuation of the function F_a , in order to maximize the value F_a ,the small iterative step value is adopted, the step value in the solving process is a constant, by the formula $x \leftarrow x + (\mu\sigma^2)\nabla F_a(x) = x - \mu\vec{\sigma}$, and the derivate the formula (6) , the lowering direction of search optimal value is obtained.

III. SPARSE REPRESENTATION CLASSIFIATION ALGOTHIIM BASED ON DEEP LEARNING

We propose a novel face recognition method based on deep network feature extraction and sparse representation classification. We extract the deep feature directly from any pose and illumination of face image, the extraction method can eliminate the differences under pose variation, and remove the intra-identity variances, while maintaining discriminativeness between identities, and the extracted features can reconstruct can recover the face image in a standard pose. Firstly, the deep network has three local connection layer and two pool layer and are alternately stacked. Each layer in different scale extract face feature. Each image outside the face region has a lot of strong response, which capture the main information of pose, illumination of face images, and some of the strong response in the area of the face capture the face structure. In the second local connection layer, a strong response in the face region has been significantly reduced, which suggests that the most of face differences are discarded and keep the face structure. The third local connection layer is reconstruction layer, which the feature is sparse and keep identity. In order to restore the standard view of face image, we use a fully connected layer. After extraction feature of face image by deep learning network, we reduce the dimension by the principal component analysis and sparse representation classification through smoothed l0 norm. The face feature were taken as vectors into the formula (4), the solution is sparse. We use smoothed sparse l0 algorithm to solve the problem, In order to solve the $\min\|\mathbf{x}\|_0 \; s.t. \; \mathbf{y}=\mathbf{Ax}$, the gradient projection method and the steepest descent method were used. The smoothed l0 algorithm can effectively reduce the recognition time. By calculating the residual error, $\gamma_i(\boldsymbol{y}) = \|\boldsymbol{y} - A\delta_i(\boldsymbol{x})\|_2, i=1,2,\cdots,K$ and using the minimum residual, $\boldsymbol{y} \in \arg\min_i \gamma_i(\boldsymbol{y})$ we can Judge the class of each test sample.

IV. EXPERIMENTS

Here we used the pose subset of the FERET face database, which includes 1400 images, including a total of 200 different people(seven of each face image), corresponding to different pose, face expression, and illumination, labeled with ba, bd, be, bf, bg, bj and bk, respectively. In the experiment, the size of each face image is 80 x 80, part of the face images are shown in figure 2 (a). Five tests with different pose angles were performed. Test 1 (pose angle is zero), image marked with ba and bj were used as training set, and images marked with bk, bg, bf, be, bd were used as test set. All experiments are performed in MATLABR2011b under 64bit Win 7 using a computer with 4G memory and 2 GHZ CPU main frequency.

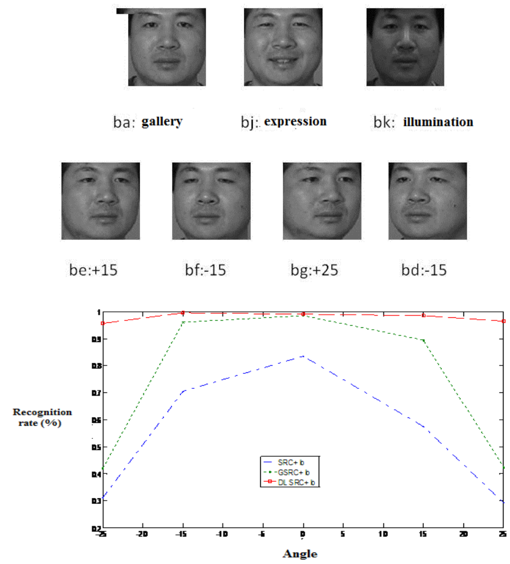


Figure 2 Part of face images on FERET pose database and recognition rates of SRC, GSRC and DLSRC (up) Part of face images of one person (down) The recongition rates of SRC, GSRC and DLSRC for different angle under normal illumination (SRC sparse classification, GSRC Gabor feature extraction and sparse classification, DLSRC feature extraction by deep learning and sparse classification)

TABLE I. THE RECOGNITION RATE(%) FOR DIFFERENT ANGLE

Angle	-25	-15	0	15	25
SRC+ l_0	31.2	70.5	83.5	57.5	29.3
GSRC+ l_0	42.1	96	98.5	89.5	42.3
D LSRC + l_0	95.6	99.6	99	98.5	96.4

It can be seen from table 1 and figure 2, the recognition rate by sparse representation classification based on deep learning is far higher than that of direct sparse classification or sparse classification after the Gabor feature extracted. when pose variation are relatively small (0-15), GSRC algorithm recognition rate are 96%, 98.5%, 89.5%, respectively, , about 20% higher than that of the SRC algorithm (70.5%, 83.5%, 57.5%), but the recognition rate of DLSRC algorithm can reach 99.6%, 99%, 98.5. The experimental results show that on FERET database, when the pose variation o face image is small, GSRC algorithm can achieve good performance, but when the pose variation is large, the performance of SRC, GSRC fell sharply, but DLSRC algorithm can still keep high recognition rate. GSRC algorithm with small pose variation has good robustness. DLSRC algorithm with big pose variation can still maintain good robustness.

TABLE II. THE RECOGNITION RATE(%) USING THE DIRECTLY EXTRACTED FEATURE

Angle	-25	-15	15	25
SRC+ l_0	42.1	60.1	67.2	37.5
GSRC+ l_0	71.2	90.1	90.8	66.5

TABLE III. THE RECOGNITION RATE(%) USING DEEP LEARNING

Angle	-25	-15	15	25
SRC+ l_o	96.1	98.9	99.1	98.5
GSRC+ l_0	92.1	99.9	99.6	96.7

In this part of the experiment, the images labeled with ba, bj, bk, were used as the training set, labeled with bg, bf, be, bd, were used as a test set. The experiments results show the face recognition rate is very low when the common feature extraction and dimensionality reduction methods, such as PCA, Gabor, are adopted and then sparse representation classification in the experiment.

The recognition rates improve sharply when the reconstructed images are input. Compare to the origin image, the recognition advantage using the reconstructed images is obvious, at the same time, the extraction Gabor, PCA feature from reconstructed image can also improve the recognition rate.

In the experiments， different methods are used in the original image and the reconstructed image, from table 2, table 3 shows that the recognition rate can increase about 30% under big pose variation when deep learning and reconstructed images are used. The algorithm can obtain relatively high recognition rate under different pose. The main reason is the reconstruction layers restore the front images. When using Gabor feature extracting and l0 norm of sparse representation classification method, under the different pose, the recognition rate using the original image are71.2%, 90.1%, 90.8%, 66.5%, the corresponding recognition rate using reconstructed image reach 92.1%, 99.1%, 99.6%, 96.7%.

TABLE IV. THE RECOGNITION TIME (S)WITH DIFFERENT METHODS FOR DIFFERENT DIMENSIONS

Dimension	30	54	130	300	540
SRC+ l_1	0.721	0.880.	1.142	1.788	5.199
GSRC+ l_1	2.693	3.326	4.473	7.928	9.131
GSRC+ l_0	0.013	0.019	0.068	0.370	2.063

Table 4 shows the recognition time required for each image when using l_1 algorithm and l_0 algorithm respectively in the cases of different dimensions after reconstruct face image. It can be seen from table 4, the sparse representation classification face recognition by using the l_1 algorithm need long time. After extracting Gabor feature of the face image, the sparse classification time by using smoothed l_0 is greatly reduced.

V. CONCLUSIONS

In this paper, a novel sparse representation face recognition algorithm based on deep network and smoothed l_0 norm is proposed. The deep learning network combined with the local and whole connection layer extract global and local information. Local connection layer of deep network use supervised CRBM, and the extracted features can recover facial image, reduce facial differences between the same class face images, while maintaining the difference between different class so as to eliminate the differences under pose variation. The l_0 norm fast sparse classification method was adopted to realize face recognition. The extracted feature from arbitrary pose face images can recover the face image in a standard pose, which can improve the performance of traditional algorithm in the case of pose variation. The l_0 norm algorithm can reduce the recognition time. The experimental results on FERET face database show that the recognition rate is increased significantly compared with the directly extracted face feature or Gabor feature algorithm. Especially when pose variation is larger, the algorithm still can maintain higher face recognition rate and reduce the recognition time. The next step is to study the video and network image using deep learning.

ACKNOWLEDGMENT

This work is supported by NNSF (No. 61072127，No. 61372193), NSF of Guangdong Province, P.R.C. (No.S2013010013311), Foundation for Distinguished Young Talents in Higher Education of Guangdong, China under grant No.2012LYM_0127, and Science Foundation for Young Teachers of Wuyi University (No. 2013zk07),and the Opening Project of Zhejiang Key Laboratory for Signal Processing (No. ZJKL_4_SP-OP2014-05).

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. Hinton. "ImageNet classification with deep convolutional neural networks", Proc. of Int. Conf. on Advances in Neural Information Processing Systems, 2012.

[2] Zhu Z, Luo P, Wang X, et al. "Deep Learning Identity-Preserving Face Space", Proc. of Int. Conf. on Computer Vision (ICCV), 2013.

[3] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. "Robust face recognition via sparse representation". IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, Vol.31, Issue 2, pp.210－227.

[4] M. Turk and A. Pentland. "Eigenfaces for recognition", Journal of Cognitive Neuroscience, 1991, Vol.3, Issue 1, pp.71-86.

[5] K. Etemad and R. Chellappa. "Discriminant analysis for recognition of human face images". J. Opt. Soc. Am. A, 1997, Vol.14, Issue.8, pp. 1724-1733.

[6] M. Bartlett, J. Movellan, and T. Sejnowski. "Face recognition by independent component analysis". IEEE Transactions on Neural Networks, 2002, Vol.13, Issue.6, pp.1450 - 1464.

[7] M. Yang, L. Zhang, D. Zhang, and S. Wang. "Relaxed Collaborative Representation for Pattern Classification", Proc. of Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 2012.

[8] Mohimani H，Zadeh M，Jutten C. "A fast approach for over complete sparse decomposition based on smothed l_0 norm". IEEE Transactions on Signal Processing，2009，Vol.57, Issue. 1, 289-301.

[9] Hyder M，Mahata K． "An improved smoothed l_0 approximation algorithm for sparse representation"． IEEE Transactions on Signal Processing，2010，Vol.58, Issue. 4, pp.2194 -2205．

[10] Amin M A and Yan H. "An empirical study on the characteristics of Gabor representations for face recognition". International Journal of Pattern Recognition and Artificial Intelligence, 2009, Vol.23, Issue.3, pp. 401-431.

[11] M. Yang and L. Zhang. "Gabor Feature based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary", Proc. of Int. Conf. on European Conference on Computer Vision , 2010.

[12] Bing Luo, Yun Zhang, Guangzhu Yu, et al. "ANN Ensembles Based Machine Vision Inspection for Solder Joints", Proc. of Int. Conf. on Control Automation, 2007, pp.3111–3115.

[13] G. B. Huang, H. Lee, and E. Learned-Miller. "Learning hierarchical representations for face verification with convolutional deep belief networks" Proc. of Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 2012.