Application of Novel Gabor-DCNN into RGB-D Face Recognition

Yuanyuan Xiao[1,2] and Xiaoyao Xie[2]
(Corresponding author: Xiaoyao Xie)
School of Computer Science and Technology, Guizhou University[1]
Huaxi District, Guiyang, China
Key Laboratory Of Information and Computing Science of Guizhou Province, Guizhou Normal University[2]
Yunyan District, Guiyang, China
(Email: xyx@gznu.edu.cn)
(Received Mar. 17, 2019; Revised and Accepted Aug. 4, 2019; First Online Dec. 13, 2019)

Abstract

In this paper, we propose a novel approach, named Gabor-DCNN, applied for face recognition technology of two modalities RGB-D images, which can extract the features through Gabor transform of images and deep convolutional neural network. Gabor transform can capture the salient visual properties with spatial frequencies and local structural features of different directions in a local area of images and enhance the object representation capability. Deep convolutional neural networks could automatically learn essential features from images compared with traditional methods. The most significant features can be obtained through the Gabor-DCNN. The final features expression of the face is performed by feature fusion of two modalities images. The experimental results indicate that the algorithm achieves much better performance than some state-of-the-art methods in terms of recognition rates on the EURECOM data set. This research provides an effective method for multiple modal face recognition under complex conditions.

Keywords: Deep Convolutional Neural Network; Face Recognition; Gabor Transform; RGB-D

1 Introduction

With the rapid development of the Internet, the technology of identity authentication based on face recognition has been extensively applied to different kinds of fields such as mobile payment, entrance guard system, and so on. Traditional algorithms of face recognition mostly use two-dimensional images, which can not effectively prevent information forgery and cope with changes such as illumination, posture, and expression. 3D images can capture more information, thus enabling higher preservation of facial detail under varying conditions [8,21]. Multidimensional spatial data acquisition equipment has been used in computer vision for many years, while the high cost limits their usage in large scale applications. With the development of sensor technology, RGB-D devices, such as Microsoft's Kinect sensor, have been widely initially used for gaming and later became a popular device for computer vision and high-quality data can be acquired easily [31]. RGB-D data obtained from binocular cameras take advantage of color images that provide appearance information of an object and Depth images [10, 17, 28] that are immune to the variations in color, illumination, rotation angle, and scale. The Depth image is a characterization of geometry, which contains accurate information related to the distance. Each pixel value of Depth image is the precise range between the sensor and the object. In recent years, a tremendous increasing number of RGB-D data is applied for different fields including object recognition, scene classification, hand gesture recognition, pose estimation, and 3D-simultaneous localization and mapping [13,23,26,27]. Depth images are also used in privacy protections [9].

In face recognition fields, there are many methods to extract features for two-dimensional images such as principal component analysis (PCA), linear discriminant analysis (LDA), which are statistic methods and can efficiently recognize targets. Many other effective approaches have been widely used such as local binary pattern (LBP), the histogram of oriented gradients (HOG), scale-invariant feature transform (SIFT) and so on. The visual features from single RGB images are not entirely handing against changes such as illuminance change. RGB-D images may potentially enhance the robustness of the feature descriptor to overcome these problems [22, 31]. There are some algorithms about the face recognition of RGB-D. Entropy and saliency are used to obtain feature maps [8], and then features are extracted by the histogram of oriented gradient (HOG), which strengthens the role of RGB images and weakens that of Depth maps. The method of 3D Local Binary Pattern (3DLBP) is used to extract fea-

tures [18], which is mainly based on histogram statistics of features spectrum. Since 2012, convolutional neural networks (CNNs) have recovered rapidly [14, 15], conquered most fields of computer vision, and are booming. CNN can be utilized as a special feature extractor to acquire stable feature representation or one of deep learning methods that combines feature extractor and classifier. The Gabor-DCNN, for the task of RGB-D face recognition, combines Gabor transform [19, 20] that is used for strengthening the texture information with the deep convolutional neural network (DCNN) [15], which is introduced for face recognition due to their excellent performance in computer vision. Through constructing the Gabor filters and a series of operations such as convolution and pooling, the correlation of images' data is fully excavated. Gabor feature maps that utilize Gabor transform to get directional properties and spatial density as the input data of the deep convolutional neural network are used to extract more salient features by fully training the network.

The proposed approach is evaluated in experiments with comparisons to some state-of-the-art methods on EURECOM data set [21]. In the domain of computer vision, while the convolutional neural network has been widely applied for face recognition and object categorization, the proposed approach could further enhance the capacity of feature representation by Gabor transform of images. As the experimental results demonstrate, the method is so robust as to facial expression variations.

The rest of the paper is organized as follows. The algorithm of Gabor-DCNN applied for RGB-D face recognition is described in Section 2. The experimental evaluations of the proposed approach with the comparisons to some state-of-the-art methods are reported in Section 3. The conclusions are drawn and future perspectives are given in Section 4.

2 The Algorithm of Gabor-DCNN

The Gabor-DCNN is presented based on RGB-D images of multi-sensor, which combines the Gabor transform [25] of images with the constructed deep convolutional neural network. Gabor transform, which can extract relevant features of images in different scales and orientations, is a robust face classification descriptor, and Gabor feature maps are widely applied as robust feature representations of images. The deep convolutional neural network as one of deep learning methods can get higher accuracy in image classification by a series of operations. The proposed algorithm can extract the most remarkable features and improve the expressive ability of the object, which is comprised of five major steps in the following subsections:

2.1 Data Preprocessing

RGB-D images obtained from binocular cameras include two types: RGB images and Depth maps, and there is a one-to-one match between them. The pixel values of the Depth image, which reflects the geometric shape of the visible surface of the object directly, represents the range between the object surface and the image capture device in the scene. The holes, existing in Depth images, which are caused by light and distance, are similar to the background, and then linear interpolation is utilized to fill the holes in three channels respectively. Viola-Jones detection [2] is carried out to acquire face regions and preserves detection frames, utilized to crop face region of the Depth map. Face detection is one of the essential techniques used in automatic processing by the computer, playing an essential role in face recognition, the aim of which is to acquire the human face information from the background contained in the digital images.

2.2 Gabor Feature Maps

Gabor feature maps can be obtained through Gabor transform, which is used to solve the problem that signals cannot be analyzed locally by Fourier transform and Gabor filters. The physical meaning of Fourier transform is to transform the gray distribution function of the image into the frequency distribution function of the image, which reflects the statistical characteristics of frequency signals, and Gabor transform is a windowed Fourier transform, the windows function of which is Gaussian function. Generally, the Gabor feature maps can be acquired in two steps:

- Constructing Gabor filters;
- Gabor transform of images by Gabor filters.

Gabor filters, which are similar to the 2D receptive field profiles of simple cells of the mammalians' visual cortex, have excellent spatial locality and directional selectivity, can grasp the local features of spatial frequency and structure in multiple directions of the image [4]. Gabor feature maps can be gotten by a set of Gabor filters, which capture the salient visual properties in images [21]. The Gabor filter is defined as follows:

$$g(x,y;\omega,\sigma,\theta) = \frac{1}{2\pi\sigma^2}\cdot exp\{-\frac{x^2+y^2}{2\sigma^2}\} \cdot exp\{j\omega(xcos\theta + ysin\theta)\}. \quad (1)$$

Here $j=\sqrt{-1}$, θ controls the orientation of the Gabor filters, $\theta = \frac{\pi\cdot m}{M}$ with $m = 0, \cdots, M-1$, and M stands for the number of orientations; ω represents frequency, $\omega = \frac{\omega_{max}}{f^n}$ with $n = 0, \cdots, N-1$, and N denotes the number of frequencies; f is the interval factor of frequency; σ is the standard deviation of the Gaussian envelope, which determines the width of Gaussian windows. The frequency ω and envelope σ control the sparsity of the Gabor feature maps.

As can be seen from Formula (1), Gabor filters are subjected to three variable: θ, ω, and σ. A specific value of Gaussian envelope: σ, will be set to 2π. Hence, the

Gabor filters can be constructed only by setting parameters: frequency and direction. We choose the maximum frequency: $\omega_{max} = \frac{\pi}{2}$, due to large-scale images usually choose smaller frequencies from these papers [3, 5, 16, 24]. Also, the Gabor feature maps with five frequencies and eight orientations are the best. The interval factor of frequency is set to $\sqrt{2}$. According to the requirement, the parameters of the Gabor filter are shown in Table 1.

Table 1: The parameters list of Gabor filters

Parameters	Values
$\omega_i(i = 1, \cdots, 5)$	$\frac{\pi}{2}, \frac{\pi}{2\sqrt{2}}, \frac{\pi}{4}, \frac{\pi}{4\sqrt{2}}, \frac{\pi}{8}$
$\theta_j(j = 1, \cdots, 8)$	$0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}$

The Gabor transform is defined as its convolutions of an image and the Gabor filters; the formal is shown as follows:

$$G(x, y) = I(x, y) * g(x, y; \omega, \sigma, \theta). \quad (2)$$

Where $I(x, y)$ stands for the images, the symbol " $*$ " denotes the convention operator, and $G(x, y)$ represents the robust feature maps obtained by Gabor transform. It is worth noting that images should be converted into gray images before Gabor transform. The features maps acquired by eight directions of Gabor filters will be combined into an 8-channel image at the same frequency, prepared for the input images of the DCNN constructed in the next subsection.

Consequently, through a series of transformation and channels fusion of images according to the rules we just mentioned, the RGB/Depth images with three channels will be converted into 8-channel images with five different frequencies respectively. The Gabor feature maps at five frequencies with eight channels can be represented as follows:

$$H_{rgb}(i) = CF_{\omega_i}(G_{\theta_j}(RGB)), \quad (3)$$
$$H_{depth}(i) = CF_{\omega_i}(G_{\theta_j}(Depth)). \quad (4)$$

Where $i \in [1, 5], j \in [1, 8]$. $CF(\cdot)$ represent the fusion of 8-channel images, which is computed by setting parameters according to Table 1. The $H_{rgb}(i)$ indicates that they are Gabor feature maps with the i-th frequency fused images of RGB images, and the $H_{depth}(i)$ indicates that they are Gabor feature maps with the i-th frequency fused images of Depth maps. Fused images aim to reduce the computational cost of feature extraction, and the most remarkable features by training the deep convolutional neural network.

2.3 Feature Extraction

In this paper, we adopt a deep convolutional neural network model to identify 8-channel fused images, which improves LeNet-5 [7] model. The LeNet-5 is a convolutional neural network including two convolution layers, two down-sampling layers, and three fully-connected layers, which is used to identify handwritten digits. The improved model, designed in Table 2, is used to recognize the fused images that are processed by Gabor transform. The improved model and Gabor transform of images can get more remarkable features.

As shown in Table 2, Conv5-6 stands for convolution layer with six filters of 5×5 size; S=2 represents stride and the value of S is 2; Max-pooling denotes that max-value is selected in pooling layer. FC represents full connection layer. The DCNN model is a six-layer deep convolutional neural network, which consists of five CS (convolution and down-sampling, CS) layers and a full connection layer. The abstract features will be extracted by convolution layers of the networks, and will be more powerful representation with the increase of convolution layers [6, 30]. The full connection layer of our network is designed to accomplish the number of object classification task and the numclass stands for the number of classes.

From Table 2, we can find that the differences between LeNet-5 and the DCNN model, which are summarized as follows. The constructed DCNN has five CS layers and a full connection layer. A CS layer stands for a convolution layer, a BN, a ReLU, and a Max-pooling layer. A ReLU, selected in the improved model, is adopted as the activation function. The activation function of the LeNet-5 model is sigmoid. Batch normalization (BN) and Dropout are used to optimize neural networks [11, 12]. The usage of BN can not only efficiently improve the training speed of the network but also enhances the generalization ability of the model. Dropout is a simple but effective regularization technique, which makes some nodes of the network not only work stochastically but also improves the generalization ability of the neural network in the training process of the neural network.

The CNN model is used to extract features. The features are extracted by the DCNN model we constructed as follows:

$$F_{rgb}(i) = C(H_{rgb}(i)), \quad (5)$$
$$F_{depth}(i) = C(H_{depth}(i)). \quad (6)$$

Where $i \in [1, 5]$, and the DCNN model is represented as $C(\cdot)$, utilized as a feature extractor. The most prominent features are selected by using a series of convolution layers and Max-pooling layers under five different frequencies. Five kinds of features obtained by extracting different 8-channel images are concatenated into a single feature vector, which represents features of a single modality. Both RGB images and Depth maps are adopted in the same way to extract features. Consequently, features of RGB-D are linked to an extended vector F as final feature expression, which is provided as the input data to a multi-class classifier. The features F are respresented as follows:

$$F = [F_{rgb}, F_{depth}]. \quad (7)$$

Table 2: Operations, dimensions and parameters list of DCNN model

Layers	Parameters
Input Layer	$256 \times 256 \times 8$
CS1 Layer	Conv5-6 (S=1), BN, ReLU, Max-pooling (S=2)
CS2 Layer	Conv5-12 (S=1), BN, ReLU, Max-pooling (S=2)
CS3 Layer	Conv6-12 (S=1), BN, ReLU, Max-pooling (S=2)
CS4 Layer	Conv5-24 (S=1), BN, ReLU, Max-pooling (S=2)
CS5 Layer	Conv5-48 (S=1), BN, ReLU, Max-pooling (S=2), Dropout
FC Layer	numclass

2.4 Pattern Classification

The selection of classifiers largely determines the efficiency of pattern recognition. Some classifiers such as Nearest Neighbour (NN), Random Decision Forests (RDFs), and Support Vector Machine (SVM) have been applied extensively into pattern classification. SVM, which is a robust supervised algorithm method of machine learning and extremely computationally accurate during probe identification, is chosen as a classifier in the proposed approach. SVM can also be applied for regression and classification, the non-linear relationship of which can get between data and features when the sample size is small and medium. Therefore, SVM is very suitable for the classification of features in this study. The category labels can be obtained by SVM.

2.5 The Computation of Accuracy

The category labels are compared with the input labels and the recognition rate (or accuracy) is calculated as follows:

$$Accuracy = \frac{Sum(L(x) == L_{input}(x))}{Numel(L_{input}(x))} \times 100\%. \quad (8)$$

Here, $L(x)$ stands for the category labels obtained by classifier; $L_{input}(x)$ represents the input label; the symbol " == " denotes object equality, and returns logical 1 (true) only where $L(x_i)$ and $L_{input}(x_i)$ are equal, the result of " $L(x) == L_{input}(x)$ " is a vector that contains 1 or 0. $Sum(\cdot)$ is utilized to calculate the sum of the vector elements, the values of which are 1. $Numel(\cdot)$ returns the number of elements, which exists in $L_{input}(x)$. To summarize, the method of Gabor-DCNN is sketched in Algorithm 1 as follows:

Algorithm 1 Gabor-DCNN

Input: The RGB-D images, I_{rgb} stands for the RGB image, I_{depth} denotes the Depth map, and the meaning of symbol " ∀ " is on behalf of all the thing.

Output: *Category Labels*, *Accuracy*

1: Begin

2: The I_{rgb} and I_{depth} are converted to grayscale images respectively at first as:

$$\begin{aligned} Ig_{-rgb} &= grayscale(I_{rgb}), \\ Ig_{-depth} &= grayscale(I_{depth}). \end{aligned}$$

3: Gabor filters are constructed according to the specific frequency, Gaussian envelope, and orientation. Details about the parameters are shown in Table 1, and a set of Gabor filters are calculated as follows by Equation (1):

$g_{ij} = \frac{1}{2\pi\sigma^2} \cdot exp\{-\frac{x^2+y^2}{2\sigma^2}\} \cdot exp\{j\omega_i(xcos\theta_j + ysin\theta_j)\}$,

$g = \forall g_{ij}$, where $i \in [1,5]$, and $j \in [1,8]$.

4: Gabor featur maps are calculated by Equation (2) for Ig_{-rgb} and Ig_{-depth} respectively as:

$$\begin{aligned} Grgb_{ij} &= Ig_{-rgb} \cdot g_{ij}, \\ Grgb &= \forall Grgb_{ij}, \\ Gdepth_{ij} &= Ig_{-depth} \cdot g_{ij}, \\ Gdepth &= \forall Gdepth_{ij}. \end{aligned}$$

Where $i \in [1,5]$ and $j \in [1,8]$.

5: The images with same frequency are combined into an 8-channel image based on Equation (3) and Equation (4) respectively, and then $Hrgb$ and $Hdepth$ stand for Gabor feature maps of RGB images and Depth maps by a series of image processing respectively:

$$\begin{aligned} H_{rgb}(i) &= CF_{\omega_i}(Grgb_{ij}), \\ Hrgb &= \forall H_{rgb}(i), \\ H_{depth}(i) &= CF_{\omega_i}(Gdepth_{ij}), \\ Hdepth &= \forall H_{depth}(i). \end{aligned}$$

Where $i \in [1,5]$ and $j \in [1,8]$.

6: The constructed DCNN model, which is a feature extractor and is represented as $C(\cdot)$, is used to extract features from fused images at a certain frequency of single modality based on Equation (5) and Equation (6), the features of two modalities images are expressed respectively as follows:

$$\begin{aligned} F_{rgb}(i) &= C(H_{rgb}(i)), \\ F_{depth}(i) &= C(H_{depth}(i)), \text{ where } i \in [1,5]. \\ F_{rgb} &= \forall F_{rgb}(i), \\ F_{depth} &= \forall F_{depth}(i), \text{ where } i \in [1,5]. \end{aligned}$$

7: The features vector F denotes the final features expression of two modalities by concatenating to reduce

the number of vectors from two to a single vector by Equation (7) as follow:

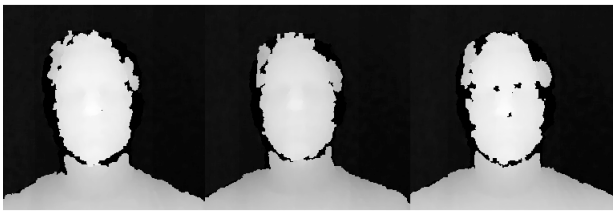
$$F = [F_{rgb}, F_{depth}].$$

8: SVM is chosen to obtain category labels and the accuracy is calculated according to Equation (8).
9: End

3 Experimental Results and Analysis

In order to validate the proposed method, we apply the proposed algorithm to RGB-D face images of EURECOM data set. Face recognition technology is a wonderfully challenging theme in the field of computer vision and pattern recognition [3, 29]. 936 images, pertaining to 52 people in EURECOM data set, are captured in the two-time series: Session 1 (S1) and Session 2 (S2), and some samples can be seen in Figure 1.

(a) RGB images



(b) Depth maps

Figure 1: The samples of the EURECOM data set

EURECOM data set has higher simulation degree with covariant variables such as pose, illumination, and occlusion. Since the EURECOM data set does not divide the testing set and the training set. It is necessary to balance the samples and divide the training set and the testing set before the experiments.

3.1 Features Extraction of Two Modalities Images

In order to test the effectiveness and the robustness of Gabor-DCNN, we will achieve the experiments for features extraction of two modalities images. Forty Gabor filters will be constructed according to the parameters in Table 1 at first, and then forty Gabor features maps of each modality can be processed by Gabor transform. The images with the same frequency will be fused into eight-channel images. Hence, ten kinds of Gabor feature maps under five different frequencies, which are used as the input of the constructed DCNN model. The recognition rates are shown in Figure 2 by fully training the DCNN model.

From Figure 2, we can find: the accuracies of the RGB images perform well, which can run at more than 96%; the recognition rates of the Depth maps are slightly lower than that of the RGB images, which are about 90%. To further investigate the performance of the proposed method, we also have tested RGB-D on other models such as LeNet-5, MT-LeNet-5, and DeepID2. LeNet-5 model is utilized to identify single-channel images such as grayscale images, we should convert two kinds of images into grayscale images as the input of the LeNet-5 model before training the model; MT-LeNet-5 means modified three-channel LeNet-5, used to achieve classification task of 3-channel images, which is different from LeNet-5 at first convolution layer; DeepID2 is also a 3-channel convolutional neural network, which is used to identify 3-channel images. The recognition rates of different network models are shown in Table 3.

Table 3: The comparison of recognition rates in different network models

Input images	Accuracy(%)		
	LeNet-5	MT-LeNet-5	DeepID2
RGB	88.78	89.42	88.46
Depth	57.37	60.58	40.38

As can be seen from the Table 3, the first column represents the input of data; the second column stands for recognition rates of LeNet-5; The third column denotes the accuracy of MT-LeNet-5; The fourth column stands for the recognition rates of DeepID2. Both RGB images and Depth maps are classified by MT-LeNet-5 and DeepID2 respectively. Grayscale images of two modalities are identified by training the LeNet-5 model. BN and Dropout are adopted to optimize the above convolutional neural network (CNN) models. From Table 3 and Figure 2, we can get better recognition rates of each modality under different frequencies by using the proposed approach than by some other methods. The most remarkable features can be extracted by both the DCNN model we constructed and Gabor transform of images, which enhances the recognition rates.

3.2 Final System and the Computation of Accuracy

The problem of features fusion based on RGB-D images will be discussed in the following and the final recognition rates will be calculated. Five kinds of features (five frequencies) of each modality are extracted and are connected in series. The features extracting from two modalities are concatenated to a vector expression as the final features, and then SVM is chosen as a classifier to achieve

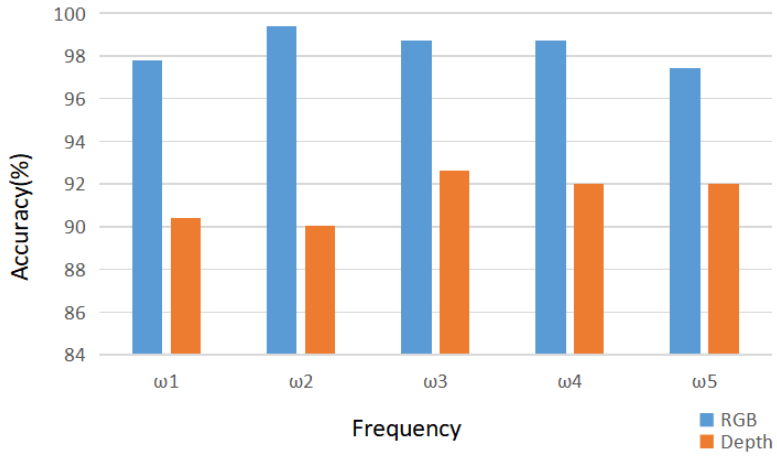


Figure 2: The recognition rates trained by the model in five frequencies

classification task.

We also compare the proposed method with RISE [8], LBP [1], LeNet-5 and DeepID2. For RISE algorithm, Viola-Jones detection is carried out at first, then the entropy maps of RGB images and Depth maps are computed respectively, and saliency maps of RGB images are calculated. The features are extracted from entropy and saliency maps by the histogram of oriented gradient (HOG), and then Random Decision Forest (RDF) is selected as a classifier to calculate recognition rates. LBP, just calculating the binary result between the center pixel and the neighbors, is used to extract features from three channels of images, and then SVM is chosen as a classifier to obtain the final result of object recognition. Both DeepID2 and LeNet-5 are utilized to extract features as extractors, features from two modalities are connected into a vector, and then SVM is chosen to achieve object recognition. The experimental results are shown in Table 4.

From Table 4, we can see that our algorithm achieves an excellent result. The identification rate using the proposed method is about 11% higher than that of the RISE algorithm, about 5% higher than that of the LBP algorithm. Since the DeepID2 and LeNet-5 cannot perform well for the cropped Depth maps by face detection, the recognition rates of two modalities images for classification task of EURECOM data set are low. The experimental results show that the proposed method performs well and is also competitive.

4 Conclusions

This paper proposes a new method, named Gabor-DCNN, applied for RGB-D face recognition. RGB-D can help solve fundamental problems due to its complementary nature of the Depth information and the visual information of the RGB. Gabor filters can magnify the changes of gray level and enhance the local features of some critical functional areas of the face such as eyes, nose, mouth, eyebrows, and so on, which are helpful to distinguish different face images. Gabor feature maps transformed by Gabor filters have excellent spatial locality and directional selectivity, which are robust to illumination and posture, and therefore have been successfully applied in face recognition. Convolutional neural network (CNN) can automatically achieve image classification and get higher accuracy than other traditional methods. The constructed DCNN model is utilized to extract features. The Gabor-DCNN, combining Gabor transform and the deep convolutional neural network can dramatically improve the classification accuracy. The experimental results on RGB-D data set validate the effectiveness of the proposed method. The proposed algorithm performs more excellently and obtains better results of object classification than some state-of-the-art.

Although we get better performance on face recognition of EURECOM data set by the proposed algorithm, the drawback of the Gabor-DCNN lies in the way of classification, which needs to extract features of two modalities RGB-D and be linked together, and then a classifier is chosen to achieve categorization. Namely, feature extractors and classifiers are separated in the proposed approach. In order to solve the problem, future work will focus on the design of the parallel neural network for face recognition, which can achieve object classification automated.

Acknowledgments

This work is supported by Natural Science Foundation of Guizhou, China under Contract LH[2014]7641, by the National Statistic Bureau of China under Contract 2016LY81. The authors thank Yu Feng and Hainan Wang for their generous assistance and valuable suggestions.

Table 4: The comparison of acccuracy in different methods on EURECOM data set

Method	The Proposed	DeepID2	LeNet-5	RISE	LBP
Accuracy	97.76	69.55	77.56	86.22	92.63

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *European Conference on Computer Vision*, pp. 469–481, 2004.

[2] A. M. Basbrain, J. Q. Gan, and A. Clark, "Accuracy enhancement of the viola-jones algorithm for thermal face detection," in *Intelligent Computing Methodologies*, pp. 71–82, 2017.

[3] Z. Baochang, S. Shiguang, C. Xilin, and G. Wen, "Histogram of Gabor phase patterns (HGPP): A novel object representation approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 57–68, 2006.

[4] Y. L. Boureau, F. Bach, Y. Lecun, and J. Ponce, "Learning mid-level features for recognition," in *Computer Vision & Pattern Recognition*, 2010. (`https://www.di.ens.fr/willow/pdfs/cvpr10c.pdf`)

[5] L. Chengjun and W. Harry, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 11, no. 4, pp. 467–476, 2002.

[6] M. Chihaoui, A. Elkefi, W. Bellil, and C. B. Amar, "Implementation of skin color selection prior to gabor filter and neural network to reduce execution time of face detection," in *International Conference on Intelligent Systems Design & Applications*, 2016. DOI: 10.1109/ISDA.2015.7489251.

[7] Y. Le Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," *Advances in Neural Information Processing Systems*, vol. 2, no. 4, pp. 396–404, 1990.

[8] G. Goswami, M. Vatsa, and R. Singh, "RGB-D face recognition with texture and attribute features," *IEEE Transactions on Information Forensics & Security*, vol. 9, no. 10, pp. 1629–1640, 2014.

[9] I. Y. H. Gu, D. P. Kumar, and Y. Yun, "Privacy-preserving fall detection in healthcare using shape and motion features from low-resolution RGB-D videos," in *International Conference Image Analysis & Recognition*, pp. 490–499, 2016.

[10] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments," *International Journal of Robotics Research*, vol. 31, no. 5, pp. 647–663, 2014.

[11] G. E. Hinton and N. Srivastava, A. Krizhevsky, I. Sutskever and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *Computer Science*, vol. 3, no. 4, pp. 212–223, 2012.

[12] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on International Conference on Machine Learning*, 2015. (`https://arxiv.org/pdf/1502.03167.pdf`)

[13] Y. Ji, A. Yamashita, and H. Asama, "RGB-D SLAM using vanishing point and door plate information in corridor environment," *Intelligent Service Robotics*, vol. 8, no. 2, pp. 105–114, 2015.

[14] A. Karpathy, G. Toderici, S. Shetty, T. Leung, and F. F. Li, "Large-scale video classification with convolutional neural networks," in *Computer Vision & Pattern Recognition*, 2014. (`https://static.googleusercontent.com/media/research.google.com/zh-TW//pubs/archive/42455.pdf`)

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems*, 2012. (`https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf`)

[16] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. V. D. Malsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture. ieee transactions on computers, 42, 300-311," *IEEE Transactions on Computers*, vol. 42, no. 3, pp. 300–311, 1993.

[17] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view RGB-D object dataset," in *IEEE International Conference on Robotics & Automation*, pp. 1817–1824, 2011.

[18] J. B. C. Neto and A. N. Marana, "Face recognition using 3DLBP method applied to depth maps obtained from kinect sensors," in *X Workshop de Visão Computacional*, 2014. (`file:///C:/Users/user/Downloads/0029.pdf`)

[19] S. Qian and D. Chen, "Discrete gabor transform," *IEEE Transactions on Signal Processing*, vol. 41, no. 7, pp. 2429–2438, 1993.

[20] S. T. H. Rizvi, G. Cabodi, P. Gusmao, and G. Francini, "Gabor filter based image representation for object classification," in *International Conference on Control*, 2016. DOI: 10.1109/CoDIT.2016.7593635.

[21] M. Rui, N. Kose, and J. L. Dugelay, "KinectFaceDB: A kinect database for face recognition," *IEEE Transactions on Systems Man & Cybernetics Systems*, vol. 44, no. 11, pp. 1534–1548, 2014.

[22] A. Schmidt, M. Fularz, M. Kraft, A. Kasiński, and M. Nowicki, "An indoor RGB-D dataset for the evaluation of robot navigation algorithms," in *International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 321–329, 2013.

[23] S. Song, S. P. Lichtenberg, and J. Xiao, "SUN RGB-D: A RGB-D scene understanding benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. DOI: 10.1109/CVPR.2015.7298655.

[24] H. Wang, B. Zhang, H. Zheng, Y. Cao, Z. Guo, and C. Qian, "The robust derivative code for object recognition," *Ksii Transactions on Internet & Information Systems*, vol. 11, no. 1, pp. 272–287, 2017.

[25] X. Wang, L. U. Youtao, S. Song, and X. Ping, "Face recognition based on Gabor wavelet transform and modular pca," *Computer Engineering & Applications*, vol. 48, no. 3, pp. 176–176, 2012.

[26] T. Whelan, M. Kaess, H. Johannsson, M. Fallon, J. J. Leonard, and J. Mcdonald, "Real-time large-scale dense RGB-D SLAM with volumetric fusion," *International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 598–626, 2015.

[27] T. Whelan, M. Kaess, J. J. Leonard, and J. Mcdonald, "Deformation-based loop closure for large scale dense RGB-D SLAM," in *IEEE / RSJ International Conference on Intelligent Robots & Systems*, 2013. `http://thomaswhelan.ie/Whelan13iros.pdf`

[28] A. Wilkowski, T. Kornuta, and W. Kasprzak, "Point-based object recognition in RGB-D images," *Intelligent Systems*, vol. 323, pp. 593–604, 2015.

[29] S. Xue, "Face database security information verification based on recognition technology," *International Journal of Network Security*, vol. 21, no. 4, pp. 601–606, 2019.

[30] H. Yao, C. Li, H. Dan, and W. Yu, "Gabor feature based convolutional neural network for object recognition in natural scene," in *International Conference on Information Science & Control Engineering*, 2016. DOI: 10.1109/ICISCE.2016.91.

[31] C. Ziyun, H. Jungong, L. Li, and S. Ling, "RGB-D datasets using microsoft kinect or similar sensors: A survey," *Multimedia Tools and Applications*, vol. 76, no. 3, pp. 4313–4355, 2017.

Biography

Yuanyuan Xiao had received M.S. in School of Electronic information Engineering from Guizhou University in 2008; she is with the School of Computer Science and Technology from Guizhou University as a Ph.D. candidate. Her research interests include machines learning, deep learning and virtual reality.

Xiaoyao Xie is a professor with Key Laboratory Of Information and Computing Science of Guizhou Province, Guizhou Normal University. He is also a member of the China Computer Federation. His current research interests include computer network, information security, machines learning and pattern classification.