

A Combined GLQP and DBN-DRF for Face Recognition in Unconstrained Environments

Hongping Hu* and Yu Yang

College of Information Science and Engineering, Hunan University, Changsha 410082, China

*Corresponding author

Abstract—This paper proposes a novel approach for accurate and robust face recognition by using Local Quantized Patterns computed from gabor-filtered images(GLQP) and Deep Belief Network ensembled dynamic random forests(DBN-DRF). GLQP is a kind of local pattern feature extractor based on gabor filters applying, it makes use of vector quantization and lookup table to let local features become more expressive without sacrificing simplicity and computational efficiency. DBN-DRF is a new deep architecture we proposed, in which dynamic random forests classifier is employed to replace inherent Softmax classifier or SVM to achieve a decent classification result at the top of network. GLQP exploits low-level local features that are used as input to DBN-DRF, which further extracts high-level abstract features for classification. Our architecture is trained and evaluated on two challenging face recognition datasets(FERET and LFW), the experiments result show our approach is competitive or better than the state of the arts.

Keywords-face recognition; local quantized patterns; gabor filters; deep belief networks; dynamic random forests

I. INTRODUCTION

Face recognition in a natural or artificial environment has been investigated extensively over the last few years, and many face recognition methods have been proposed [1][2]. Despite of this methods have achieved reasonably good performance in controlled conditions, their performance is still unsatisfactory in unconstrained conditions where diverse environments such as varying lighting, expression and pose affect the recognition performance heavily. Hence, how to extract robust and discriminative features to enlarge the inter-personal margins and reduce the intra-personal variations simultaneously remains a central and challenging problem in the field.

Advances in visual representation and classification method have been the major source of progress in the field of face recognition. Currently, local patterns features(such as Local Binary Patterns(LBP)[1], Local Quantized Patterns(LQP)[2], Gabor filters[1],etc.) are being used quite successfully in [3],[4],[5] for face representation because of their ability to robustly encode local information for each image pixel as well as their computational simplicity. Nevertheless, the above methods are doped with subjective factors, and the extracted features depend on manual selection. In 2006, the depth model (such as Convolutional Neural Network(CNN), Deep Auto-encoder Networks(DAN), Deep Belief Network(DBN), etc.) emerged as an automatic feature learning tool without manual intervention. While they have achieved reasonably good performance in different applications such as digital recognition and human detection [7][8][9][10][11][12], the depth network

will learn the unfavorable feature description when regards the original pixels as input, since these pixels are affected by changing lighting, posture, and so on.. Moreover, in the real situation, the training sample set is usually small and not sufficient to adjust the network weights, the problem also weakens the performance of the network.

In this paper, we propose an effective way to learn expressive and robust features by combing GLQP and DBN-DRF for face recognition in unconstrained environments, in order to overcome the shortcomings of the above methods as well as improve the recognition accuracy. For face recognition, different face regions usually have different local structures and it is desirable to learn more region-specific features for face representation. A possible way to achieve this goal is using GLQP to learn features for different face local regions individually. The additional use of Gabor filters before LQP can significantly improve the robustness to changes in lighting conditions[15], the variety of scales and orientations of garbor filters can also increase the datasets. However, different face local regions usually share some related information in feature representation and individual feature learning ignores this characteristic. To exploit shared information among different face local regions, we apply Deep Belief Network(DBN) to automatically and comprehensively process all the local features from GLQP, where both the relationship between different face local regions and position-specific information are exploited simultaneously. After obtaining the high-level abstract features from the DBN, we perform dynamic random forests as classifier to compare visual representation for classification and recognition, by the reason that we experimentally validate the dynamic random forests achieves the best accuracy relative to the inherent Softmax and SVM.

II. RELATED WORK

Most of the previous studies use low-level local features to characterize facial traits. T.Ojala introduced the original LBP in 1994,this theoretically very simple and efficient multiresolution operator labels the pixels of an image by thresholding the 3×3 neighborhood of each pixel with the center pixel value and considers the result as a binary number. [5]determined edges of key points in color image for optimization real time human face detection with LBP. [6]designed an improved HOG-LBP algorithm to profile the features of depth face information, and then used SVM for face classification. Despite the high coding efficiency of LBP, histograms of these local features increases exponentially when the processing circle diameter is increased. [2]introduced Local Quantized Patterns(LQP) by modifying the LBP with a learned coding strategy to solve the above problems.

[3]proposed a discriminant face descriptor, called DS-LQP, to increase the recognition performance for depth image via a bank of SVM classifiers. [4]proposed a new method named Local Vector Quantization Pattern (LVQP) that based on the generation of the difference vector of a central pixel with its neighbors. Although low-level region-specific features have achieved encouraging results on visual representation, they ignored the sharing information between different image regions.

In order to extract high-level abstract features automatically, not just low-level local features extracted under artificial intervention, several alternatives have been proposed, e.g.[7]used Convolution Neural Network (CNN) with noisy labels, [8]developed Cascaded Deep Auto-Encoder Network (CDAN)approach,[9][10]adopted Deep Belief Networks(DBN), [11][12]used Deep Boltzmann Machine(DBM), [13]built multiple deep ConvNets to learn high-level global features of face and trained RBM for classification. Although the above methods of extracting features are joint and automatic, the extracted features are sensitive to illumination, posture, and other interference in unconstrained environments. In addition, these deep models ignore the local structure of the image and learn some unfavorable feature expression when take the original pixel as input directly.

III. METHODOLOGY

Our face recognition framework contains two stages: first, local quantized patterns based on gabor image are used to characterize the facial position-specific traits, then the deep belief networks embedded dynamic random forests is used to extract global related features for recognition —c.f. Figure I. The following subsections give relevant details on each of these stages.

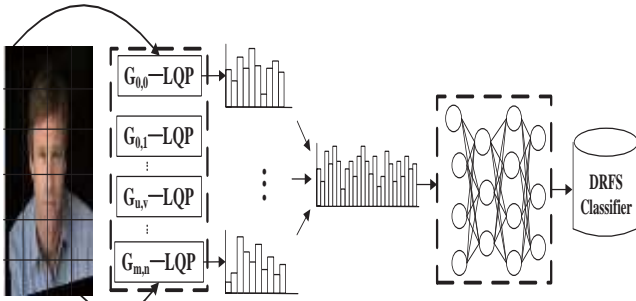


FIGURE I. THE SCHEME FOR THE PROPOSED FRAMEWORK

A. Region-Specific Local Features Extracting using GLQP

1) Local Quantized Patterns(LQP). LQP[2] is a generalized form of local patterns that use large local neighbourhoods and deeper quantization with domain-adaptive vector quantization. It uses binary/ternary codes which produced from the comparisons of surrounding pixel values with the central pixel to span a discrete space e.g. a local pattern that involves 24 binary pixel comparisons generates $2^{24}=17$ million distinct codes, while for 16 ternary generates $3^{16}=43$ million distinct codes, and divides them into groups by clustering[14]. LQP allows a much wider range of neighbourhoods geometries including: horizontal(H), vertical(V), diagonal(D) and antidiagonal(A) strips of pixels; combinations of these like horizontal-vertical(HV) and diagonal-antidiagonal(DA); and traditional

circular and disk-shaped regions. Geometries will be described by notation such as DA2 7, where DA indicates the neighbourhood shape, the subscript indicates the neighbourhood diameter and the superscript denotes the quantization level. The diverse and flexible topologies of LQP are helpful to exploit more information available in different image region than other local patterns e.g. LBP or LTP.

In this paper, we use the HVDA2 7 layout (c.f. Figure II) to represent facial local traits. HVDA denotes the type of geometry of horizontal-vertical-diagonal-antidiagonal, the subscript indicates the neighbourhood diameter is 7 pixel and the superscript indicates the quantization level is binary code.

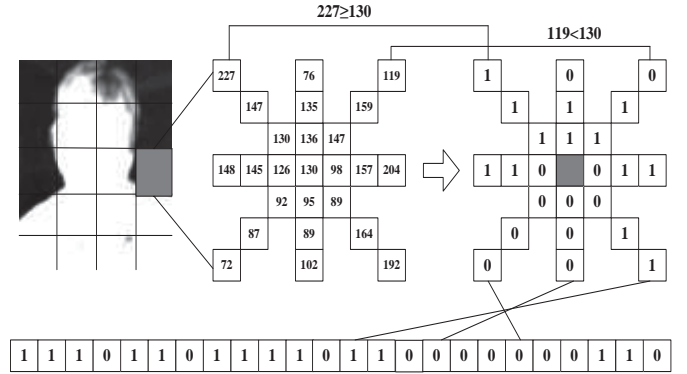


FIGURE II. FACE REPRESENTATION USING LQP

2) GLQP(Gabor LQP). Gabor filters, acclaimed to be optimally localized in the spatial and frequency domains, permit the local structures of input images to be extracted with respect to the tuned scale and orientations, and have been successfully used for face recognition[1][15] due to their robustness in unconstrained environments. They have been used as a complementary feature set to local pattern features as well as the pre-processing stage of feature extraction. In GLQP, LQP features are computed from gabor filtered image obtained by convolving the image with multi-scale multi-orientation gabor kernels. In this paper, 40 different gabor kernels that span 5 different scales and 8 different orientations over the range 0 to 2π , in practice of face image preprocessing, are defined as follows:

$$\psi_{u,v} = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{(-\|k_{u,v}\|^2 \|z\|^2 / 2\sigma^2)} [e^{ik_{u,v}z} - e^{-\sigma^2/2}], \quad (1)$$

Where $u \in \{0,1,2,3,4\}$ determines the scale, $v \in \{0,1,2,3,4,5,6,7\}$ controls the orientation, $z=(x,y)$, $\sigma=2\pi$.

3) Implementation Details. In our implementation of the proposed approach, all face images from different databases are cropped and scaled to size 90×150 pixels, then Gabor filters are applied to normalized image. In order to fully characterize the local position of the face image, we divide the image into κ non-overlapping pixel cells and build cell-level histograms using HVDA2 7. In this paper $\kappa=9 \times 15=135$. After HVDA2 7 applying, these 24-bit codes form huge array of binary sequences, we consider only those codes that occur more than N times to discard outliers, N set to 10 in the experiments. The resulting

codes are stored in a lookup table, GLQP builds the table by mapping all the codes to the closest center using K-Means method[14]. By this means, codes are divided into λ clusters form a dictionary, in this paper, $\lambda=200$, our dictionary is presented in table I. We then concatenate all theses cell-level histograms from all cells of every gabor image to form the GLQP feature of the sample. The final histogram has $135 \times 200 \times 40=1,080,000$ bins for $\kappa=9 \times 15=135$ pixel cells, $\lambda=200$ clusters, and 40 different gabor kernels. The connected features are mapped to a low-dimensional feature space using WPCA as the final local expression of the face image.

TABLE I. DICTIONARY FOR $\Lambda=200$

24-bit binary code	Number of cluster
0000 0000 0000 0000 0000 0000	1
0000 0000 0000 0000 0000 0001	1
.... ..	
0000 0000 0000 1110 0111 0001	84
0000 0000 0000 1110 0111 0010	84
0000 0000 0001 1110 0111 0011	96
.... ..	
1011 0101 0110 1100 1010 1000	127
1011 0101 0110 1100 1010 1001	127
1011 0101 0110 1100 1010 1011	153
.... ..	
1111 1111 1111 1111 1111 1111	200

B. Global Related Features Extracting and Face Recognition using DBN-DRF

For extracting high-level global related features without artificial intervention, we use DBN-DRF based on previous low-level local features learned by GLQP. In the following we provide details on the complete process.

1) Structure of DBN-DRF. DBN-DRF is a neural network stacked by restricted boltzmann machines (RBM) and dynamic random forests(DRF), where each RBM is atop of its lower RBM, DRF is a classifier at the top of the network. In our study, the input of lowest RBM is region-specific feature vectors learned by GLQP, while the output of DRF is labels of global abstract feature vectors of face. RBMs in the middle layers between bottom RBM and top DRF extract features in data that comes from their below RBM. We set the number of network layers to 4, including an input layer, two hidden layers, and an output layer, as illustrated in Figure III..

Each RBM is a two-layer network in which stochastic input vectors are connected to stochastic output feature detectors using symmetrically weighted connections. The input layer is visible and each unit corresponds to a probability like value in a certain dimension. The feature detectors correspond to hidden units. According to energy theory, there exist an energy which is relevant to a joint configuration (v, h) of the visible and hidden units given by

$$E(v, h; \theta) = \sum_{i=1}^m \sum_{j=1}^n W_{ij} v_i h_j - \sum_{i=1}^m b_i v_i - \sum_{j=1}^n a_j h_j, \quad (2)$$

where v_i and h_j are the states of unit i in visible input v and unit j in hidden feature h , a_j and b_i are their biases, w_{ij} represents the weight between them. The joint distribution over the visible and hidden vectors is:

$$P(v, h; \theta) = \frac{1}{Z(\theta)} \exp(-E(v, h; \theta)), \quad (3)$$

$$Z(\theta) = \sum_{h, v} \exp(-E(v, h; \theta)). \quad (4)$$

For a DBN with l -layer hidden units, the joint distribution of its visible and hidden layers can be expressed as:

$$P(v, h^{(1)}, \dots, h^{(l)}) = P(v | h^{(1)}) P(h^{(1)} | h^{(2)}) \dots P(h^{(l-1)} | h^{(l)}), \quad (5)$$

where v represents the visible layer, $h^{(m)} (m=1, 2, \dots, l)$ refers to the m th hidden layer. The hidden layer m and $m+1$ meet the relationship as follows:

$$P(h^{(m)} | h^{(m+1)}) = \prod_i P(h_i^{(m)} | h^{(m+1)}). \quad (6)$$

In this paper, we propose to combine GLQP with DBN-DRF to extract discriminative information for face recognition, the joint distribution of the deep network in our model is:

$$P(Q, h^{(1)}, h^{(2)}, \dots, h^{(l)}) = P(Q | h^{(1)}) P(h^{(1)} | h^{(2)}) \dots P(h^{(l-1)} | h^{(l)}), \quad (7)$$

where Q denotes region-specific features learned by GLQP, while $h(1), h(2), \dots, h(l)$ is the advanced features learned by DBN in different levels of input data Q .

2) Network Training and Testing. Following Hinton's work in [17][18], we train and test our network in three steps. The first step is pre-training process of each RBM, which is a greedy layer-wise unsupervised procedure, pre-training one layer at a time in a greedy way. Initialize the visible layer with training data, then calculate the hidden layer with the conditional distribution, the visible layer is reconstructed from the hidden layer according to the conditional distribution symmetrically. After learning one layer of feature detectors, the values of hidden units can be treated as input data to the second layer for extracting features. The first layer of feature detectors then become the visible units for learning the next RBM. We set the learning rate to 0.05 and set the number of iterations to be 500. The second step we combine all pretrained RBMs together to fine-tune resultant DBN with labeled data using BP(back propagation)algorithm, we set the number of iterations to be 50. The last step is done by adding DRF to the top of the network, training the whole network by stochastic gradient descent on the cross-entropy with respect to the target classification and then testing the face recognition accuracy in FERET[19] and LFW [20] datasets. The DRF algorithm[21] is based on a sequential

procedure that builds an ensemble of random trees by making each of them dependent on the previous ones. It exploits the same idea with adaptative resampling process of boosting and combines it with the randomization processes used in “classical” RF. In this paper, the number of trees to be built is 500.

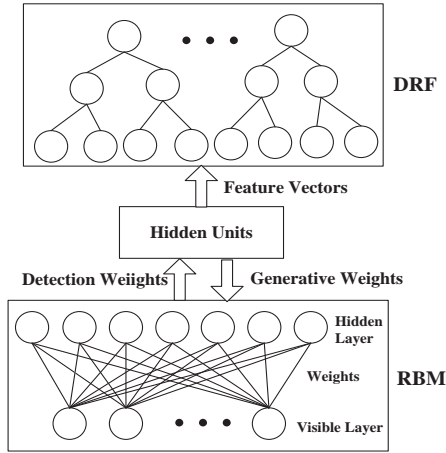


FIGURE III. ILLUSTRATION OF DBN-DRF STRUCTURE

IV. EXPERIMENTS

We carry out our experiments on FERET[19] database and LFW[20] database to show the effectiveness of our approach. The face images in FERET and LFW present significant variations in lighting, expression and pose. The number of neuron nodes in DBN is set to 10800, 5000, 2500 respectively, and the comparative methods are carried out with similar configuration. All experiments were conducted with MATLAB R2015b on a PC with 3.40GHz Intel Core i5-4210CPU, Windows7OS and 4 GB main memory. We will explain the experiment for each database in details as follows.

A. Evaluation on the FERET Dataset

The FERET dataset contains 14,051 facial images of 1196 different individuals with resolution 384×256 pixels. The complete dataset is partitioned into two disjoint sets: gallery and probe. Gallery has labeled images and is used only for training, while the probe is used for testing, which is further subdivided into four categories: fb, fc, dup1 and dup2. In our experiments, we use fa as gallery set and take fb, fc, dup1, and dup2 as probe sets. The experimental results of our algorithm comparing with other popular algorithms are shown in Table II.

It can be seen obviously that our method shows its superiority compared with other methods, getting respectively 59.1%, 34.3%, 6.8%, and 4.3% higher mean accuracy than LBP, LQP, DBN, and LBP+DBN, and outperforming them in all four separate subsets. From the last three rows, we can know that using the DRF classifier at the top of the network leads to significant improvement in the accuracy, improving the mean accuracy of SVM and Softmax classifiers at least 0.3%. We record the following two observations from the above results: (i)the flexible structure for position-specific feature learning and robustness to variations in unconstrained environments enable GLQP to encode local information of face image successfully; (ii)the DBN-DRF takes the GLQP features as input rather than

the original pixel, which helps the depth network to learn more favorable feature representation.

TABLE II. THE EXPERIMENTAL RESULTS ON FERET DATABASE

Algorithm	Testing set				
	Fb	Fc	Dup-I	Dup-II	Average
LBP(OpenCV)	76.7%	21.1%	38.1%	20.9%	39.2%
LQP(OpenCV)	98.6%	45.4%	61.6%	50.4%	64.0%
DBN	91.6%	92.3%	91.8%	90.2%	91.5%
LBP+DBN	98.8%	92.8%	92.9%	91.5%	94.0%
GLQP+DBN+SVM	98.4%	97.2%	96.9%	97.0%	97.4%
GLQP+DBN+Softmax	99.1%	97.7%	97.5%	97.6%	98.0%
GLQP+DBN-DRF	99.4%	98.1%	97.8%	98.0%	98.3%

B. Evaluation on the LFW Dataset

Experiment has also been conducted in LFW dataset. LFW consists of 13233 face images of 5749 different individuals of different ethnicity, gender, age, etc., of which 4069 people have only one image, 1369 people have 2~5 images, the other 311 people have more than 6 images. Most of the images are color images, a few for the gray scale, with resolution 250×250 . LFW is a typical face database for face recognition under unrestricted conditions, reflecting various changes of face images under real conditions, such as pose, lighting, expression, background, etc., so the performance testing of the face recognition program is closer to natural environment.

In our experiment, we select those who have the number of images greater than or equal to 20 as the experimental object, which include 62 people altogether 3023 images. we evaluate our method with two different protocols: (i) different number of training samples and (ii) different number of experimental subjects that randomly selected 10, 20, 30 as three groups of subjects from the identified 62 people, and 5 images are randomly selected as training samples, the rest are test samples for each person. The verification performance are tabulated in Table III and Table IV, respectively.

Because it is difficult to collect a large number of label training samples under realistic conditions, so it is critical to the algorithm for achieving the desired performance under the conditions of less training samples. Table III shows the fluctuation of the recognition rate of each algorithm with the increase of the number of training samples. As can be seen, our method fluctuates by at most 0.6%, while the contrast methods fluctuate by at least 1.6%, at least 1% higher than our method, indicating that our method has stronger stability under unrestricted conditions. On the other hand, as shown in Table IV, when the number of classification categories increases, our algorithm still maintains the highest recognition accuracy and does not reduce the accuracy with the increase of categories of classification, which shows that our method outperforms other algorithms in terms of robustness and universality under unrestricted environments.

TABLE III. ACCURACY ON DIFFERENT QUANTITY OF SAMPLES

Algorithm	Number of training samples		
	5	10	15
SVM	37.1%	45.2%	56.5%
DBN	83.9%	85.5%	87.1%
LQP+SVM	82.3%	84.0%	87.1%
LBP+DBN	87.1%	88.7%	90.5%
GLQP+ DBN-DRF	95.3%	95.9%	96.4%

TABLE IV. ACCURACY ON DIFFERENT QUANTITY OF CATEGORIES

Algorithm	Number of categories		
	10	20	30
SVM	50.0%	55.0%	53.3%
DBN	40.0%	40.0%	36.6%
LQP+SVM	80.0%	85.0%	86.7%
LBP+DBN	100.0%	90.0%	90.0%
GLQP+ DBN-DRF	100.0%	95.0%	93.3%

V. CONCLUSION

In this paper, we revisit some popular algorithms including the basis of some local feature learning-based LBP and LQP, as well as some typical machine learning-based DBN and CNN. We analyze the limitations of these methods and propose a novel approach combined GLQP and DBN-DRF to solve the problems of these methods in terms of visual representation and classification under unrestricted conditions. Different from the traditional methods, our algorithm puts position-specific features extracted by GLQP into DBN visible layer, and uses DRF for classification at the top of deep architecture, which is under the idea that learn multiple different yet related local features automatically and comprehensively, and classify them effectively. The encouraging results verify the effectiveness of our method, as well as the rationality of our idea. Owing to the stability and universality, our method is very suitable for real-time application.

ACKNOWLEDGMENTS

This work is supported by Natural Science Foundation of China (no.61672216 and no. 61673162).

REFERENCES

- [1] X Tan and B Triggs, "Fusing Gabor and LBP Feature Sets for Kernel-Based Face Recognition," *Analysis and Modeling of Faces and Gesture, DBLP*, vol.4778, pp.235-249,2007.
- [2] Hussain S U and Triggs B, "Visual recognition using local quantized patterns," *British Machine Vision Conference,Berlin*,vol.s3-2, pp.99.1-99.11,2012.
- [3] Tomás Mantecón, Carlos R.del-Blanco,Fernando Jaureguizar and Narciso García, "Access control based on visual face recognition using Depth Spatiograms of Local Quantied Patterns," *IEEE International Conference on Consumer Electronics*, vol.117-119, pp.530-531,2015.
- [4] Pan Z, Fan H, and Li Z, "Texture Classification Using Local Pattern Based on Vector Quantization," *IEEE Transactions on Image Processing*,Vol.24,pp.5379-5388,2015.
- [5] Islomov Shahboz Zokir Ugli and Bobokulov Mirzohid Mirsalim Ugli, "Optimization detection of smiling and opening eyes in face with algorithm LBP," *International Conference on Information Science and Communications Technologies (ICISCT)*, pp.1-4,2016.
- [6] Li T, Hou W, Fei L, Lei Y, and Chen X, "Face Detection based-on Depth Information Using HOG-LBP," *Sixth International Conference on Instrumentation & Measurement, Computer, Communication and Control* ,pp.779-784,2016.

- [7] Wu X, He R, Sun Z, and Tan T, "A Light CNN for Deep Face Representation with Noisy Labels," *Computer Science*,2016.
- [8] Weng R, Lu J, Tan Y P, and Zhou J, "Learning Cascaded Deep Auto-Encoder Networks for Face Alignment," *IEEE Transactions on Multimedia*, vol.18, pp.1-1,2016.
- [9] Huang G B,Lee H, and Learned-Miller E, "Learning hierarchical representations for face verification with convolutional deep belief networks," *IEEE Conference on Computer Vision and Pattern Recognition*,vol.157, pp.2518-2525,2012.
- [10] Mazdak Fatahi, Mahmood Ahmadi, Arash Ahmadi, Mahyar Shahsavari,and Philippe Devienne, "Towards an spiking deep belief network for face recognitionapplication," *International Conference on Computer and Knowledge Engineering (ICCCKE)*pp.153-158,2016.
- [11] S Wang, HE Menghua, Z Gao, HE Shan, and JI Qiang, " Emotion recognition from thermal Infrared images using deep Boltzmann machine," *FCS*, vol. 8, pp. 609-618,2014.
- [12] Gan Q, Wu C, Wang S,and Ji Q,"Posed and spontaneous facial expression differentiation using deep Boltzmann machines," *International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp.643-648,2015.
- [13] Y Sun, X Wang, and X Tang,"Hybrid deep learning for face verification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.38, 2013, pp.1997-2009.
- [14] Wu X, Kumar V, Ross Quinlan J, Ghosh J,Yang Q, et al., "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol.14,pp.1-37,2008.
- [15] Yi j, Su F," Gabor Orientation Histogram for Face Representation and Recognition,"*Lecture Notes in Electrical Engineering*, vol.278, pp.35-43,2014.
- [16] Hinton G E, " Learning multiple layers of representation," *Trends in Cognitive Sciences*,vol.11,pp.428-434,2007.
- [17] Hinton G E, and Salakhutdinov R R, " Reducing the Dimensionality of Data with Neural Networks," *Science*, vol.313,pp.504-7,2006.
- [18] G.E.Hinton,Osindero S, and Teh Y W, "A fast learning algorithm for deep belief nets," *Neural computation*, vol.18,pp.1527-1554,2006.
- [19] Phillips P G,Wechsler H,Huang J,and Rauss P J, "The FERET database and evaluation procedure for face-recognition algorithms," *Transactions on Image and Vision Computing*, vol.16,pp.295-306,1998.
- [20] Huang G B, Mattar M, Berg T, and Miller E L, "Labeled faces in the wild:A database for studying face recognition in unconstrained environments,"*Technical Report, University of Massachusetts,Amherst*, Month 2007.
- [21] Bernard S, Adam S, Heutte L, "Dynamic Random Forests," *Pattern Recognition Letters*, vol.33,pp.1580-1586,2012.