Proceedings of the 2015
IEEE Conference on Robotics and Biomimetics
Zhuhai, China, December 6-9, 2015

An Improved eLBPH Method for Facial Identity Recognition: Expression-Specific Weighted Local Binary Pattern Histogram

Xuanyang Xi[1], Zhengke Qin[2], Shuguang Ding[3] and Hong Qiao[4], *Senior Member, IEEE*

***Abstract*— Face perception is one of the most important tasks in robot vision especially for service robots. The spatially enhanced local binary pattern histogram (eLBPH) method has been proved to be effective for facial image representation and analysis, but the expression factor isn't considered and the region-dividing method is rough. In this paper, inspired by the biological mechanism of human memory and face perception, we improve the eLBPH and propose a new method, expression-specific weighted local binary pattern histogram (EWLBPH). Accordingly, the new method introduces a semantic division process and an extended modulation process into the classical eLBPH. What's more, for the facial expression recognition, we propose a novel method which utilizes the convolutional deep belief network (CDBN) to extract discriminative information and represent them effectively. Finally, through experiments we verify the rationality and effectiveness of the improvement and two psychophysical findings.**

I. INTRODUCTION

The local binary pattern (LBP) was firstly proposed in [1] and then used as a kind of spatial feature for texture analyses. The operator has a tolerance to monotonic gray-scale and illumination variations, and has been demonstrated to be highly discriminative [2]. Ahonen et al. utilize it as a face descriptor for identity recognition [3]. They divide a face image into several regions from which local binary pattern histograms are extracted, and concatenate these basic histograms into an eLBPH which represents both the appearance and the spatial topology of divided facial regions. Additionally, they refer to the fact that some facial components play more important roles in face recognition than others and make a corresponding transformation, namely weighted eLBPH. However, they don't consider the factor of facial expression. The contribution of each facial component varies under different expressions to the identity recognition. Meanwhile, their region-dividing method is rough and each face region doesn't have explicit semantic meaning.

Current facial identity recognition approaches with considering facial expression can be divided into four categories. The first category concentrates on identifying the facial expression and transferring an expressive image into a neutral one used for final face recognition [4], [5], [6]. The second category focuses on using the optical flow (OF) which can reflect the motions of faces [7]. The third one focuses on calculating the similarity between the probe image and the gallery by dividing the image into separated regions, with larger weight on the regions which are less dependent on facial expressions, such as nose [8], [9], [10]. The last one focuses on using some elastic 3D face models to simulate the facial expression [11]. However, these methods only work for 3D face recognition instead of traditional 2D face recognition.

In this paper, we improve the classical eLBPH method for facial identity recognition and propose a new method, namely expression-specific weighted local binary pattern histogram. The new method mimics the biological mechanism of human memory and face perception that facial expression recognition and facial identity recognition are processed respectively and interact to process further. Accordingly, the proposed method introduces an extended modulation process into the eLBPH where the expression recognition modulates the identity recognition. For the sake of effective modulation, the method also introduces a semantic division process using which the divided regions are approximately component-based and own coarse but explicit semantic meaning. What's more, for the facial expression recognition, we propose a novel method which utilizes the CDBN to complete the feature learning and feature selection synchronously. The CDBN can spontaneously locate the places which contain the most discriminative information used for expression recognition and represent them effectively.

The main contributions of this paper are summarized below.

1) The proposed method mimics the biological mechanism of face perception and memory in human brain.
2) We introduce a semantic division process and an extended modulation process, where the facial expression recognition modulates the facial identity recognition, into the classical eLBPH. The rationality and effectiveness of the improvement are confirmed by experiments.
3) For facial expression recognition, we propose a novel method which utilizes the CDBN to extract discriminative information and represent them effectively. Its effectiveness is also verified by experiments.
4) We confirm two psychophysical findings that the contribution of each facial component varies under different expressions to the facial identity recognition, and the most

*This work was supported in part by the National Natural Science Foundation of China under Grant 61033011, and in part by the National Science and Technology Major Project of China under grant number 2014ZX04013011.

[1]X. Xi is a Ph.D. candidate with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190. xixuanyang2012@ia.ac.cn

[2]Z. Qin is with the Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190. qinzhengke2012@ia.ac.cn

[3]S. Ding is with the Institute of Applied Mathematics, AMSS, Chinese Academy of Sciences, Beijing 100190. s.g.ding@amss.ac.cn

[4]H. Qiao is with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190. hong.qiao@ia.ac.cn

978-1-4673-9675-2/15/$31.00 © 2015 IEEE

discriminative information for facial expression recognition is mainly located around eyes, nose and mouth.

II. BIOLOGICAL EVIDENCES

In this section, we explain the corresponding biological evidences which inspire the proposed method. Especially the two visual pathways in face-perception and their interaction are the core evidences to motivate the framework of the proposed method. The implementation of the final identity recognition stage refers to the memory mechanism of human.

A. Two Distinct Visual Pathways in Face-Perception

For face perception, Bruce and Young [12] firstly proposed a multi-route model processing seven distinct types of information derived from faces. Later, Haxby et al. [13] proposed a compatible model for the visual perception of faces, which contains two functionally and neurologically distinct pathways: one for perception of changeable aspects (including posterior superior temporal sulcus (pSTS)) and the other for perception of invariant aspects (including the fusiform face area (FFA)). The changeable aspects refer to expression, eye gaze, lip movement, etc. And the invariant aspects refer to personal identity, name, biographical information, etc.

B. Interaction between Expression and Identity Recognition

Even though the perception of facial expression and facial identity are processed via two distinct pathways, they still have some interaction. Neuroimaging studies find that the occipital face area (OFA) involved in the early perception of facial features has a feed-forward projection to the FFA and pSTS. That is, the early perception has a feed-forward projection to changeable-aspect perception and invariant-aspect perception. Recent studies [14] reported that the pSTS has a functional connection with FFA. Baseler et al. [15] found that distinct neural pathways involved in expression and identity interact to process the changeable features of a face.

C. The Memory Mechanism of Human

In psychology, memory consists of three processing stages: storing, encoding and recall. Storing is the process to place newly acquired information into brain. Encoding refers to the process of converting the information into a construct to make the retrieval easier and conscious thinking be recalled. Recall refers to the retrieval of information acquired previously. It starts with a search and retrieval process and then continues with a decision or recognition process to choose correct information from retrieved results [16], which is well known as Austin Simonson theory.

III. EWLBPH FOR FACIAL IDENTITY RECOGNITION

Inspired by above biological evidences, we propose the EWLBPH method which consists of three serial parts stated below. The relationship between the three parts is just like the corresponding biological mechanism explained in section II-B. That is, the first part has a feed-forward projection to the latter two parts, and the third part is modulated by the second part.

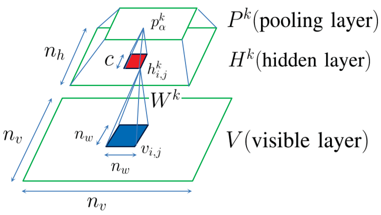


Fig. 1: A single layer CRBM with max-pooling. We only show one feature map for simplicity.

A. Locating Facial Key Components

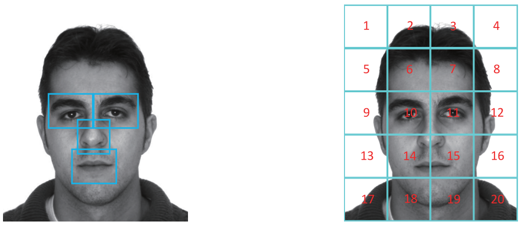
The function of this part is to estimate the center locations of facial key components: right eye, left eye, nose and mouth. This part is based on the biological evidence that a early face perception happens before the expression and identity recognition, and the early perception have a feed-forward projection to the expression and identity recognition.

As the current performance of locating facial landmarks is acceptable [17], [18] and is not our key focus, we just use the method in [19] which is a derivative of active shape model (ASM). It utilizes a simplified scale-invariant feature transform (SIFT) descriptor instead of classical 1D gradient profiles, and utilizes multivariate adaptive regression splines for matching. More details can be seen in [19].

B. Facial Expression Recognition

The function of this part is to recognize the facial expression based on the biological evidence that one visual pathway in face perception is responsible for processing changeable-aspect information including facial expression. In this paper, we utilize the CDBN to complete the feature learning and feature selection synchronously for expression recognition. The CDBN can spontaneously locate the places which contain the most discriminative information used for expression recognition and represent them effectively. In detail, four CDBNs are utilized and given names as CDBN-RE, CDBN-LE, CDBN-N and CDBN-M. They are respectively responsible for processing face patches from the four facial organs: right eye, left eye, nose and mouth. After separately extracting discriminative features for expression recognition, a concatenated feature is passed to a support vector machine (SVM) classifier to complete the expression recognition.

1) Convolutional Deep Belief Network: Here, the CDBN is a combination of two layers of convolutional restricted boltzmann machine (CRBM) [20] which is a two-layer, bipartite and undirected probabilistic graphical model. Often, a CRBM is usually followed by a pooling layer to prevent from over-fitting. A simplified structure of this single layer CRBM with max-pooling is illustrated in Fig. 1. It includes three layers: visible layer, hidden layer and pooling layer. The visible layer V is an $n_v \times n_v$ dimensional array of real-valued units. The filter weight W^k is $n_w \times n_w$ dimensional. The hidden layer H^k consists of K groups of $n_h \times n_h$ ($n_h = n_v - n_w + 1$) dimensional arrays. Each hidden unit is locally connected to visible units and the filter weights are



(a) (b)

Fig. 2: (a) For expression recognition, a face image is roughly divided into four semantic patches. (b) For identity recognition, a face image is divided into twenty semantic square regions. Better view in the electronic form.

shared among all locations within each group in hidden layer. In addition, each group has a bias b_k and all visible units have a shared bias c. The pooling layer P^k utilizes probabilistic max-pooling. The energy function of the CRBM with real-valued inputs is defined as:

$$E(\mathbf{v},\mathbf{h}) = -\sum_{k=1}^{K}\sum_{i,j=1}^{n_h}\sum_{r,s=1}^{n_w} h_{ij}^k W_{rs}^k v_{i+r-1,j+s-1} - \sum_{k=1}^{K} b_k \sum_{i,j=1}^{n_h} h_{ij}^k - c\sum_{i,j=1}^{n_v} v_{ij} + \frac{1}{2}\sum_{i,j=1}^{n_v} v_{ij}^2 \tag{1}$$

The CRBM can be trained using the contrastive divergence approximation [21] just as how RBM is done. By stacking CRBMs one to another, where the outputs of the previous CRBM are regarded as the inputs of the followed CRBM, a multi-layer CDBN is formulated.

2) Facial Expression Recognition with CDBN: Firstly, based on the estimation of the center locations of facial key components in section III-A, a face image $face$ is roughly divided into four semantic patches $patch_i$ $(i = 1,2,3,4)$ which cover the four facial organs as shown in Fig. 2a. The four face patches being as inputs are respectively send to different CDBNs but having the same structure. Fig. 3 shows the simplified structure of CDBN-LE which is responsible for processing the patches $patch_2$ from the left eye. At the top of each CDBN, the output is a low-dimensional array which characters a face region for expression recognition. For the sake of computing, the low-dimensional array is transformed into a vector x_i. Then a concatenated feature $X = [x_1, x_2, x_3, x_4]$ is passed to a SVM classifier which gives the expression recognition result $Y_E = k$ $(k = 1,2,\ldots,N_e)$ where N_e is the total number of face expressions.

C. Facial Identity Recognition Under the Modulation of Expression

The function of this part is to extract the EWLBPH and utilize it for the facial identity recognition under the modulation of facial expression. This part is based on the biological evidences that one visual pathway in face perception is responsible for processing invariant-aspect information including identity, and two distinct neural pathways involved in expression and identity recognition interact to process the

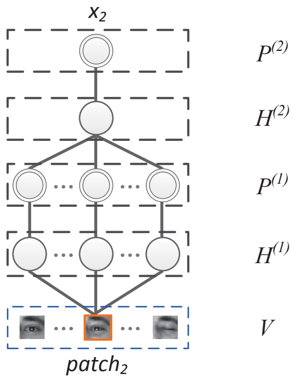


Fig. 3: The simplified structure of CDBN-LE.

changeable features of a face. The procedure consists of three stages as below.

1) The Preprocessing: Taking into account human face characteristics, a face image $face$ is divided into N_r semantic square regions $region_i$ $(i = 1,2,\ldots,N_r)$ based on the estimation of the center locations of facial key components in section III-A. As shown in Fig. 2b, we make the total number of face regions $N_r = 20$ in this paper. In order to make each region be enclosed with explicit semantic meaning, the dividing is person-specific and need to comply with two simple rules: a) the length of the side of each square is equal to the distance between two eyes l_{eye}; b) two eye centers are regarded as the centers of $region_{10}$ and $region_{11}$. Thus the regions are approximately component-based and own coarse but explicit semantic meaning. For example, the $region_6$ is the right forehead, the $region_{11}$ is the left eye and the $region_{16}$ is the left cheek.

2) Extracting the EWLBPH: After the dividing, we calculate a local binary pattern histogram for each face region. We carry out the LBP operating on each face region $region_i$, which means thresholding a neighborhood of each pixel in $region_i$ with the center pixel value and assigning the binary result to each pixel in the output labeled image $f_i(x,y)$. It can be illustrated as

$$f_i(x,y) = \sum_p 2^p s[v_p(x,y) - v_c(x,y)] \tag{2}$$

where v_c is the center pixel value, v_p is the adjacent pixel value, s is a threshold function and p is relevant to the neighborhood type. Then a histogram H_i of each labeled image $f_i(x,y)$ is calculated as

$$H_i^j = \sum_{x,y} I\{f_i(x,y) = j\}, H_i^j \in H_i, j = 0,...,n-1 \tag{3}$$

where n is the number of labels in the LBP operating.

As the expression recognition result $Y_E = k$ has been obtained in part B, each histogram H_i is enclosed with a weight $\omega_{i,k}$ which represents the contribution of the ith face region to the identity recognition under the kth facial expression. The $\omega_{i,k}$ can be confirmed with a trick in which the identity recognition accuracy just via the ith region on the training set only consisting of samples of the kth expression is assigned to $\omega_{i,k}$. All the region weights $\omega_{i,k}$ compose a modulation weight set Ω_k ($\omega_{i,k} \in \Omega_k$).

3) Final Facial Identity Recognition: The final facial identity recognition refers to the recall mechanism in human brain which starts with a retrieval process, and then chooses correct information from retrieved results in a decision process. Practically, the final facial identity recognition can be completed as

$$Y_I = \underset{j\in\{1,2,\ldots,N_p\}}{\arg\max} \left(\sum_{i=1}^{N_r} \omega_{i,k} P_{i,j} | Y_E = k\right) \tag{4}$$

where N_p is the number of memorized people and $P_{i,j}$ represents the similarity with the jth memorized people under neutral expression only with respect to the ith face region. Note that the similarity is just compared with neutral samples. It can reduce the storage and complexity of the method effectively. The meaning of (4) is equivalent to the decision process in human recall.

Particularly $P_{i,j}$ can be estimated via a extending support vector machine with pairwise coupling which can give probability estimation instead of label prediction for multi-class classification. The method is illustrated as

$$\begin{aligned} p_i &= \sum_{j:j\neq i} \left(\frac{p_i + p_j}{k-1}\right) r_{ij} \\ \text{s. t.} \quad & \sum_{i=1}^{k} p_i = 1, p \geq 0, \forall i. \end{aligned} \tag{5}$$

where r_{ij} is the estimation of $p\left(y = i \,|y = i \text{ or } j, x\right)$. More details are in [22]. The meaning of (5) is equivalent to the retrieval process in human recall.

The procedure of facial identity recognition with EWLBPH is summarized in Algorithm 1.

Algorithm 1: Facial Identity Recognition with EWLBPH

Input : A face image $face$ to be recognized, the expression label $Y_E = k$ and corresponding modulation weight set Ω_k.
Output: The facial identity Y_I.

Divide a face image $face$ into N_r semantic square regions $region_i$ as described in section III-C.1.
for $i = 1; i \leq N_r; i++$ **do**
  Compute an LBP image $f_i(x, y)$ for each face region $region_i$ using (2).
  Compute a histogram H_i for each labeled image $f_i(x, y)$ using (3).
  for $j = 1; j \leq N_p; j++$ **do**
    Compute the similarity $P_{i,j}$ with the jth memorized people just with respect to the ith face region using (5).
  end for
end for
Computing total similarities under the modulation of expression, and then obtain identity recognition result Y_I using (4).

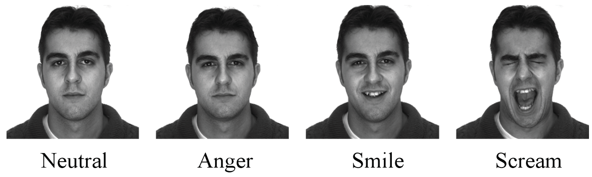


Fig. 4: Four expressions of the same person in the AR database.

IV. EXPERIMENTS AND ANALYSIS

To demonstrate the rationality and effectiveness of the proposed method for facial identity recognition, we conduct experiments on the AR database [23]. Especially as our region dividing method is component-based, we study the interference of glasses through comparison experiments on different subsets. In addition, we implement several experiments under different experiment conditions to explore the characteristics of the proposed method. We also perform comparison experiments with classical identity recognition algorithms. This section includes four parts.

A. Performance of the Proposed Method on the AR Database

The AR database consists of over 4000 color images for 126 people including 70 men and 56 women. The same picture is taken in two sessions, so hereafter the pictures are obviously divided into a training set and a test set according to the shoot session. Each person is taken with four expressions: neutral, anger, smile and scream, as shown in Fig. 4. We can find that the neutral expression and the anger expression are almost the same. In fact, the rule is suitable for almost all the people in the database. Meanwhile, as the lack of neutral samples, we approximately regard the anger samples as the neutral samples hereafter. Thus, each person has three neutral samples for the training process.

What's more, for the purpose of researching the interference of glasses, we divide the AR database into two subsets: AR-A and AR-B, according to whether a person wears glasses or not. The AR-A subset consists of 80 people wearing no glasses. After eliminating some people whose images are damaged or lost, the rest of the AR database belong to the AR-B. In both subsets, each person has three different expression samples for test.

In order to reduce parameters, the four CDBNs share the same model parameters. The first-layer CRBM has $K_1 = 10$ feature maps while the second-layer CRBM has $K_2 = 1$ feature map. In all CRBMs, the width of a convolutional window n_w is 8 and the width of a pooling window c is 2. Additionally, we utilize classical circular local binary pattern and sample 8 points on a circle of radius 1.

Firstly, we carry out the EWLBPH-based facial identity recognition on the AR-A and AR-B respectively and the results are given in Table I. The recognition accuracy on the

TABLE I: Recognition accuracies on different databases.

Database	Expression	Identity
AR-A	96.67%	94.58%
AR-B	95.56%	98.89%
AR-A + AR-B	95.76%	95.15%

AR-B is higher than that on the AR-A. It can't be regarded as a powerful proof for the interference of glasses. Maybe the factor that the number of people in the AR-B is less leads to the success of AR-B.

Then we combine the AR-A and AR-B and carry out the same experiment further. This time, the recognition accuracy is almost the same as the one on the AR-A. Through analyzing these results, we infer that the EWLBPH-based identity recognition method is robust to glasses, and hereafter we don't need considering the interference of glasses.

Similarly, from the Table I we can draw conclusion that the CDBN-based expression recognition method seems also robust to glasses. In addition, the good performance confirms the effectiveness of the expression recognition method, and verifies a psychological finding that the most discriminative information for expression recognition is mainly located around eyes, nose and mouth [24].

B. Visualization of Modulation Weight Sets

We visualize the three modulation weight sets obtained on the AR database in Fig. 5 where the spatial relation is consistent with that in Fig. 2b. The darker a region is, the more a region contributes to the identity recognition. Through analyzing the three colormaps, we can make some conclusions as below.

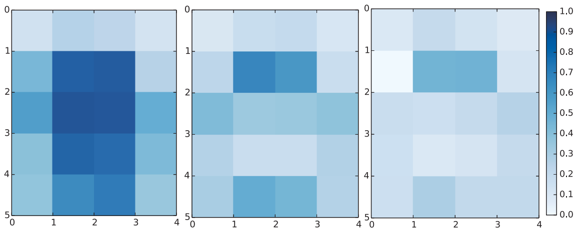
1) Under the neutral expression, the main facial components: forehead, eyes, nose and mouth, contain more discriminative information than others for identity recognition.
2) Face transformation can corrupt the discriminative ability of facial components for identity recognition. And the greater a face transformation is (e.g., from neutral to scream vs from neutral to smile), the more serious a corruption is.
3) For identity recognition, the ability of forehead to resist the corruption caused by face transformation is stronger than that of eyes, nose and mouth.
4) As the semantic division obeying two simple rules is coarse, the four corners of face images contain discriminative information more or less. Especially the two bottom corners fairly contain the texture information about clothes which is useful for the identity recognition.

In total, the visualization of modulation weight sets demonstrates the rationality of the proposed method for facial identity recognition, and verifies a psychophysical finding that the contribution of each facial component varies under different expressions to the facial identity recognition.

C. Characteristics of the Proposed Method

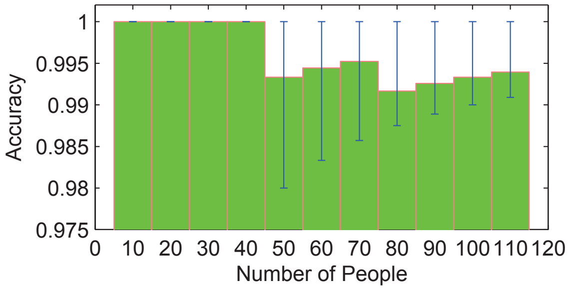
To further study the characteristics of the proposed method for facial identity recognition, we separately carry out identity recognition tasks on neutral, smile and scream samples in the AR database when the number of people varies. Fig. 6 plots the three experiment results from which we have following three conclusions.

1) The recognition accuracy declines slowly as the number of people rises, and reaches a relatively stable state at the eighty people.
2) The stability of recognition performance is in inverse

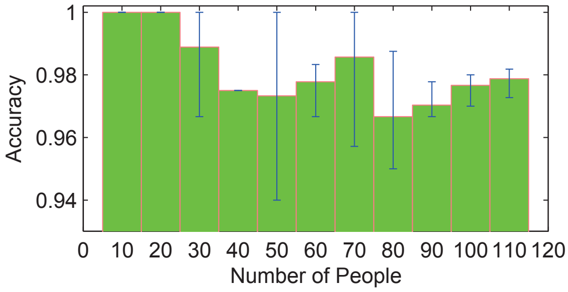


(a) Neutral (b) Smile (c) Scream

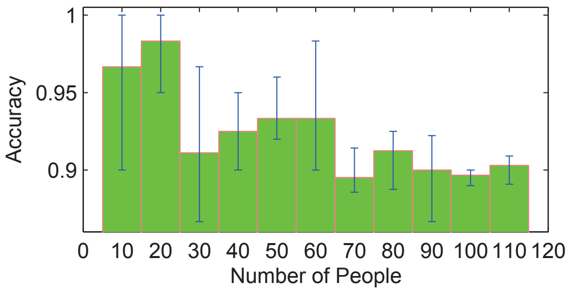
Fig. 5: The visualization of the three modulation weight sets.



(a) The identity recogniton accuracy on the neutral samples



(b) The identity recogniton accuracy on the smile samples



(c) The identity recogniton accuracy on the scream samples

Fig. 6: The facial identity recognition accuracies separately on different expression samples.

proportion to the degree of face transformation. For example, the stability of recognition performance on the neutral samples is higher than that on scream samples. That's because face transformation degrades the reliability of the texture-based feature for identity recognition.

3) At a fixed number of people, the recognition accuracy is in inverse proportion to the degree of face transformation. For example, the recognition accuracy on the scream samples is always lower than that on smile samples. This can be explained with the same reason as above.

D. Comparison with Other Classical Methods

To verify the effectiveness of the improvement, we conduct a comparison experiment with classical weighted eLBPH and eigenface on the AR database. Particularly, the weighted eLBPH and EWLBPH adopt the same parameters for the basic LBP operating. The identity recognition performance of the three methods is shown in Fig. 7. From the overall trend of the performance, the EWLBPH is better than other two methods. It verifies the success of our improvement, namely introducing a semantic division process and an extended modulation process. And it also demonstrates the effectiveness of the proposed method for facial identity recognition.

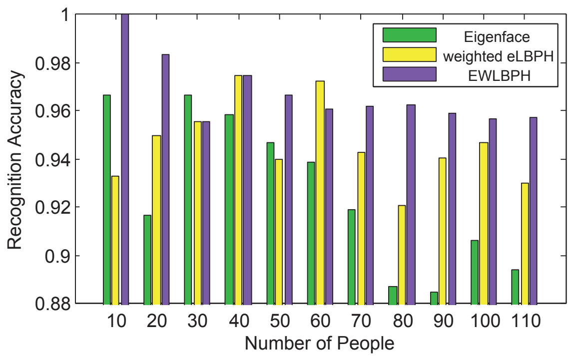


Fig. 7: The performance comparison with weighted eLBPH and eigenface. Each person has three different expression samples, neutral, smile and scream for test. Electronic form with discriminative colors for better view.

V. CONCLUSIONS

In this paper, we introduce a semantic division process and an extended expression-based modulation process into the eLBPH for facial identity recognition. For facial expression recognition, we propose a novel method which utilizes the CDBN to extract discriminative information and represent them effectively. Through above experiments, we verify the rationality and effectiveness of the EWLBPH-based facial identity recognition method and the CDBN-based facial expression recognition method. What's more, we also confirm two psychophysical findings.

REFERENCES

[1] D. HE and L. Wang, "Texture unit, texture spectrum, and texture analysis," *IEEE Trans. Geoscience and Remote Sensing*, vol. 28, no. 4, pp. 509–512, Jul. 1990.

[2] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[3] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* , vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[4] X. Gao, Y. Su, X. Li, and D. Tao, "A review of active appearance models," *IEEE Trans. Syst., Man, Cybern. C*, vol. 40, no. 2, pp. 145–158, Mar. 2010.

[5] H.-S. Lee and D. Kim, "Expression-invariant face recognition by facial expression transformations," *Pattern recognition letters*, vol. 29, no. 13, pp. 1797–1805, 2008.

[6] C. Petpairote and S. Madarasmi, "Face recognition improvement by converting expression faces to neutral faces," in *2013 13th Int'l Symp. Communications and Information Technologies (ISCIT)*, Sept. 2013, pp. 439–444.

[7] C.-K. Hsieh, S.-H. Lai, and Y.-C. Chen, "Expression-invariant face recognition with constrained optical flow warping," *IEEE Trans. Multimedia*, vol. 11, no. 4, pp. 600–610, 2009.

[8] H. Boughrara, L. Chen, C. Ben Amar, and M. Chtourou, "Face recognition under varying facial expression based on perceived facial images and local feature matching," in *2012 Int'l Conf. Information Technology and e-Services (ICITeS)*, Mar. 2012, pp. 1–6.

[9] W.-Y. Lin, M.-Y. Chen, K. Widder, Y. H. Hu, and N. Boston, "Fusion of multiple facial regions for expression-invariant face recognition," in *IEEE 9th Workshop Multimedia Signal Processing (MMSP)*, Oct. 2007, pp. 426–429.

[10] S. Fernandes and G. Bala, "A comparative study on ica and lpp based face recognition under varying illuminations and facial expressions," in *2013 Int'l Conf. Signal Processing Image Processing Pattern Recognition (ICSIPR)*, Feb. 2013, pp. 122–126.

[11] K. Chang, W. Bowyer, and P. Flynn, "Multiple nose region matching for 3d face recognition under varying facial expression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1695–1700, Oct. 2006.

[12] V. Bruce and A. W. Young, "Understanding face recognition," *British Journal of Psychology*, vol. 77, pp. 305–327, 1986.

[13] J. Haxby, E. Hoffman, and M. Gobbini, "The distributed human neural system for face perception," *Trends in Cognitive Sciences*, vol. 4, no. 6, pp. 223–233, Jun. 2000.

[14] N. B. Turk-Browne, S. V. Norman-Haignere, and G. McCarthy, "Face-specific resting functional connectivity between the fusiform gyrus and posterior superior temporal sulcus," *Frontiers in Human Neuroscience*, vol. 4, Sep. 2010.

[15] H. A. Baseler, R. J. Harris, A. W. Young, and T. J. Andrews, "Neural Responses to Expression and Gaze in the Posterior Superior Temporal Sulcus Interact with Facial Identity," *Cerebral Cortex*, vol. 24, no. 3, pp. 737–744, Mar. 2014.

[16] M. J. Watkins and J. M. Gardiner, "An appreciation of generate-recognize theory of recall," *Journal of Verbal Learning and Verbal Behavior*, vol. 18, no. 6, pp. 687 – 704, 1979.

[17] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *2012 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2012, pp. 2879–2886.

[18] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial Landmark Detection by Deep Multi-task Learning," in *13th European Conf. Computer Vision (ECCV)*, vol. 8694, Sep. 2014, pp. 94–108.

[19] S. Milborrow and F. Nicolls, "Active Shape Models with SIFT Descriptors and MARS," *VISAPP*, 2014.

[20] H. Qiao, X. Xi, Y. Li, W. Wu, and F. Li, "Biologically inspired visual model with preliminary cognition and active attention adjustment," *Cybernetics, IEEE Transactions on*, vol. 45, no. 11, pp. 2612–2624, Nov 2015.

[21] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[22] T. Wu, C. Lin, and R. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, Aug. 2004.

[23] A. Martinez and R. Benavente, "The ar face database," *CVC Tech. Report 24*, 1998.

[24] J. Cohn and A. Zlochower, "A computerized analysis of facial expression: Feasibility of automated discrimination," *American Psychological Society*, vol. 2, p. 6, 1995.