FACIAL EXPRESSION RECOGNITION USING SVM CLASSIFICATION ON MIC-MACRO PATTERNS

Housam Khalifa Bashier Babiker[1], Randy Goebel[2], and Irene Cheng[3]

Department of Computing Science, University of Alberta, Canada

***Abstract*— The identification of facial expressions is a fundamental topic in the area of human computer interaction and pattern recognition. The research has gained significant attention in recent years. However many challenges still exist. This is because an individual might display different expressions at different times for the same mood. Expressions can also be influenced by health. Our proposed framework aims to capture unique information related to expressions from salient patches. We extract representative feature patterns at both micro and macro levels within a pixel-patch, and use a support vector machine (SVM) classifier to label expressions. Our experimental results using the Japanese facial expression (JAFEE) and Cohn-Kanade (CK) datasets achieve high recognition rate and efficient computation time, outperforming existing work.**

Index Terms—: Facial expression, salient features, support vector machine, micro-pattern, macro-pattern, dual-view expression.

I. INTRODUCTION

Facial expressions can be seen as the changes on the face due to the responses of our brain to social communication, emotions and intentions [11]. Technological advances makes it possible for machines to identify expressions in a variety of applications, e.g., sociable robotics, interactive games, and assisted rehabilitation. Three main steps are needed to label expressions: (1) face detection, (2) feature extraction and (3) emotion classification.

Face detection methods are used to localize a face in an image [17]. *Feature extraction* methods are applied to extract unique representations and are divided into three general approaches: geometric, hybrid and appearance feature extraction [28,24,21,16]. The geometric approach relies on the geometry of the face such as shape, location and distance between landmarks. The hybrid approach combines geometric and appearance features [6]. However, these methods are computationally expensive, which makes them not feasible in many circumstances, e.g., on smart-phones, which have limited capacity. The appearance approach encodes facial signals. A variety of algorithms have been proposed in this area, e.g., Gabor wavelet [5,6], which captures the appearance variation of an expression. To address the complexity and memory allocation issues [35], the local binary pattern algorithm (LBP) was proposed [2,18,23]. LBP encodes features by thresholding a 3 x 3 pixel neighbourhood using the value of the center-pixel, and then converts the output to a binary pattern.

Motivated by graphical models, the Symmetric Local Graph Structure (SLGS) was introduced as a new operator [1]. Different from LBP, SLGS uses all the pixels in the graph to obtain the binary pattern, instead of thresholding with the center-pixel. However, SLGS is not robust against noise and cannot detect important fine detail.

Many *emotion classification* algorithms are used for labeling facial expressions, e.g., support vector machines (SVM) [22,34,17], k-nearest neighbour classifier (KNN)[22], and techniques based on linear programming [26,8]. Dimensionality reduction steps are also applied by researchers to reduce the feature space before classification [19,14,13].

The biologically-based method [4] proposed to follow the phenomenon that humans can identify expressions from both near and far distances. The method requires two normalized facial images at two scales, i.e., large and small faces, with eight fiducials points (salient regions): forehead, two eyebrows, two eyes, two cheeks, and mouth. LBP is extracted from local regions along with local phase quanitizaton (LPQ) [22] to overcome the problem of scaling, noise and occlusion. However, noise reduction is not guaranteed along multiple scales. Locating the fiducials points and extracting their features make the framework computationally expensive and not feasible when computational capacity is limited, e.g. on mobile devices. Furthermore, not all salient regions chosen carry useful information.

Therefore, despite recent advancements, for computers to label expressions is still challenging because facial expression is a dynamic entity. A person's emotion can change over time and vary under different circumstances [4]; a person can have different expressions depending on his or her physical state, mood or mental health [7]. Our contribution lies in the formulation of a new salient-region based localization and feature extraction framework using micro- and macro-patterns within a pixel patch. We successfully reduce the number of salient regions (fiducial points) to only two, leading to both better recognition rate and lower computational cost.

The rest of this paper is organized as follows. Section II introduces our proposed framework. Section III describes the experiments and finally, Section IV concludes our findings and discusses future work.

II. The proposed framework

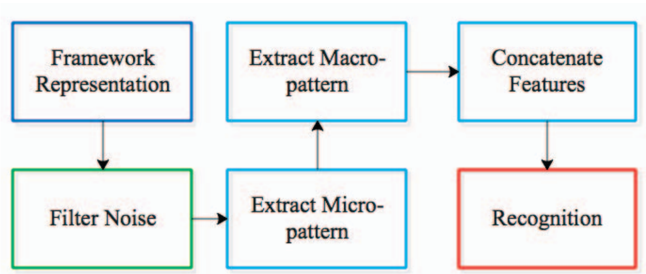

Fig. 1: Block diagram of our proposed framework.

Expression features can be represented in histograms. However, a histogram depicting a facial expression from a single-view does not have the capacity to embed spatial information, because it only has 256-bins to represent the whole image. This problem can be resolved by dividing the facial patches in a way that the algorithm can capture better spatial information.

We introduce a dual-view framework using two faces simulating a far-view and near-view. The far view provides the global structure while the near view provides the local detail. Each view is divided into four salient patches to capture the changes of expression at both micro and macro levels. We focus on the two patches of the eyes because eyes carry key information related to expression. The micro- and macro-patterns are extracted from each patch and concatenated together to yield a feature vector, which contains the unique expression representation for a given image patch I. This feature vector is forwarded to the support vector machine either for training or classification. Our framework is shown in Fig. 1.

Our graphical design aims to extract both global (far-view) and local (near-view) features of a facial image defined as macro- and micro-patterns within a pixel-patch. The macro-pattern encodes the global information while the micro-pattern encodes the local fine details (Fig. 2).

A. Macro-pattern encoding

In our framework, unique features of different expressions are encoded in a directed graph. Each edge connecting two nodes encodes a down-flow or up-flow using a binary weight. The flow is computed by subtracting the end node value from the start node value. The resulted eight binary weights form a feature string, which is then converted to a decimal feature value.

In order to reduce noise, we first filtered each pixel $I(x, y)$ using a nonlinear filter ϕ. Since this filter is applied to the entire salient patch and in order to reduce

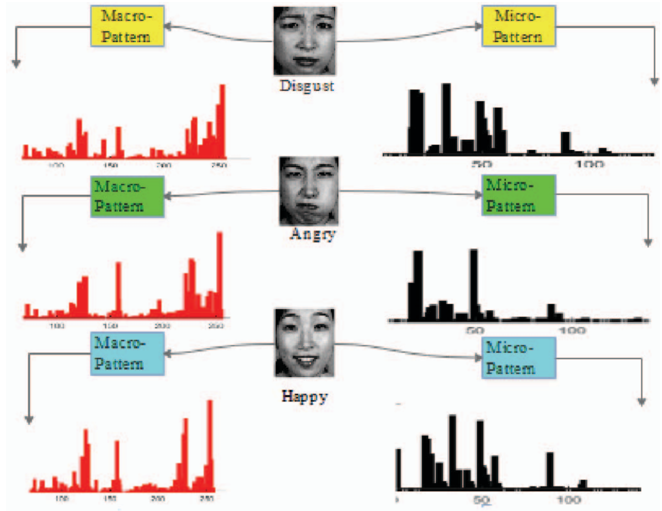

Fig. 2: Examples of extracted micro- and macro-patterns of different expressions in a given view using our proposed approach. It can be seen that there is a distinction between micro and macro patterns of the same emotion. Also, the patterns are different depending on the emotion as shown by the histograms.

the computational time, We approximate a filtered patch $I^{'}$ as follows:

$$I^{'} = I - \mu_I \tag{1}$$

where I is the noisy patch and μ_f denotes the mean over the whole patch defined as follows:

$$\mu_I = \frac{1}{N} \sum_{n=0}^{N} (I_n) \tag{2}$$

where N is the number of pixels in the patch, and I_n denotes the n^{th} expression pixel.

To obtain a macro-representation of a given target pixel $I^{'}(x, y)$ (colored dark-blue on the left in Fig. 3), we first define a neighborhood A as follows:

$$A = \sum_{i=x+1}^{x+W} \sum_{j=y}^{y+L-1} I^{'}_{ij} \tag{3}$$

where W is the x dimension, L is the y dimension , S defines the starting node, and R defines the radius with the target expression pixel as the center. Graph nodes are connected by directed edges.

$$W = (4 * R) + 1, L = (2 * R) + 1, S = R + 1 \tag{4}$$

We observe that in general facial expressions propagate more horizontally than vertically, e.g., smiling, and thus $W > L$ in the neighborhood A.

Inside A, every pixel is either a graph node or non-graph node, and the size of A depends on R. A directed edge is encoded 1 if the start node has a value greater than or equal to the end node, and is encoded 0 otherwise. The node values are gray-scale values in the range $0 - 255$. An example of the encoding is given in Fig. 4.

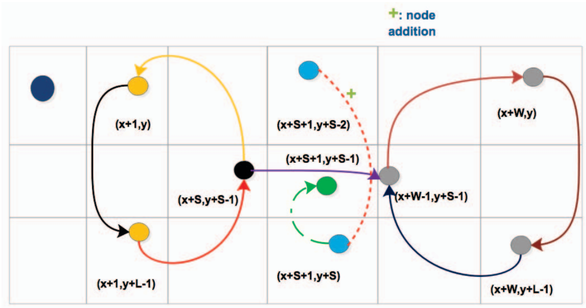

Fig. 3: Macro-pattern neighborhood when $R = 1$. Empty squares represent pixels not designated as graph nodes.

For the pixel $I^{'}(x, y)$ (dark-blue node), its marco-pattern is computed by following the arrows starting from pixel $(x+S, y+S-1)$ (black node). A weight of 1 is obtained by finding the flow between the two connected nodes i.e. $(x + S, y + S - 1), (x + 1, y)$. We repeat this process for every pair of nodes in the directed graph except for the nodes labeled light-blue and green. This 8^{th} weight is obtained by finding the difference between the green node and the sum of the two light-blue nodes as shown in Fig. 4. The last step is to convert the weight string into a 8-bit gray value.

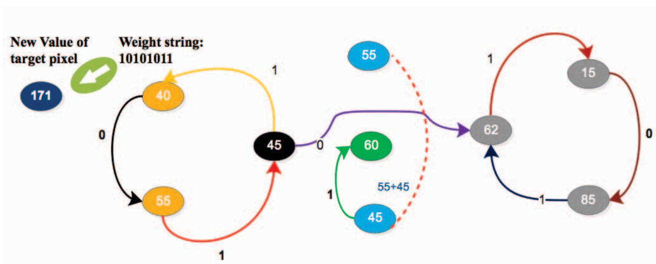

Fig. 4: Illustration of the macro-pattern encoding process. The target pixel is encoded 171 derived from the weight string 10101011.

Note that as R increases, the number of non-graph nodes also increases. The starting black node moves further away from the target pixel $I^{'}(x, y)$. We observe that pixels close to the target pixel often display similarity. On the other hand, pixels too far away may display characteristics irrelevant to the target pixel. By increasing R, we are able to obtain an optimal value and collect neighborhood information to enhance the feature pattern, leading to best recognition rate and lower computational cost. Our finding is reported in the Experimental Section.

B. Micro-pattern encoding

Similar to macro-pattern extraction, a noise filter ϕ is first applied (Eq. 1). The same radius R defined in Eq. 3 and neighborhood A are used. The encoding process is defined as follows:

$$f_{dt} = \sum_{n=0}^{7} f_{d,n} - \mu_D \tag{5}$$

where f_{dt} is the new pixel representation, $f_{d,n}$ represents a facial pixel and μ_D denotes the local mean of the yellow nodes in the neighborhood (Fig. 5). We use μ instead of median because the values of the nodes are closer to the mean and not the median. The micro-pattern weight W

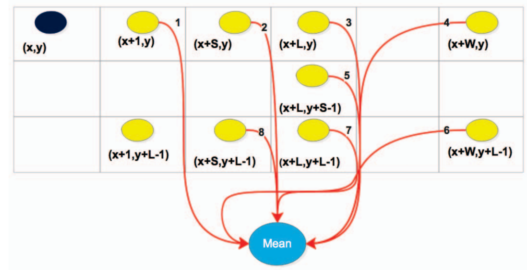

Fig. 5: Illustration of our micro-pattern structure encoding process when $R = 1$.

of an expression pixel $I^{'}(x, y)$ is obtained by forwarding the yellow nodes to the mean function below:

$$W_i = \begin{cases} 1, & \text{if } k - \mu_D > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where k denotes the value of a yellow node. The weight W is used to approximate the 8-bit gray-value of a target pixel (x, y) (colored dark-blue). In Fig. 5, the number on each red line indicates the order of calculating the 8-bit feature string.

Micro-pattern captures the local changes of the pixel I(x,y). The pattern is obtained by thresholding the graph nodes in Fig.5 using the local mean. Macro-pattern captures the global structure obtained by considering the global relation of the graph nodes in Fig.4.

III. Experiments and Analysis

A. Support vector machine (SVM) for expression classification

SVM classification is applied on macro- and micro-features extracted. Every feature h is normalized to unit variance as follows:

$$\begin{aligned} H_t^{'} &= \sum_{i=0}^{N} (h_{ti} - \mu_{h_{ti}} / \sigma_{hti})) \\ H_e^{'} &= \sum_{i=0}^{N} (h_{ei} - \mu_{h_{ei}} / \sigma_{hei})) \end{aligned} \tag{7}$$

where $H_t^{'}$ and $H_e^{'}$ are the normalized features for training and testing, h_{ti} and h_{ei} represent feature vectors for training and testing respectively, μ is the mean and σ is the standard deviation.

We employ the commonly used SVM classifier [5] to recognize expressions. We performed our experiments on two widely used datasets, the Japanese Female Facial Expression (JAFFE) [20] and the Cohn-Kanade (CK) database [15].

B. Experiments on (JAFEE) database

The JAFEE database has seven expressions and the 213 pictures are obtained from 10 females with 3 to 4 facial images for the same expression.

1) *Experiments setup:* Two approaches were used to examine our proposed method, i.e., 10-fold cross validation and leave-one-out. For fair comparison, we followed the setup proposed in [4]. 210 images were used for cross-validation to balance the size of each fold, while all the images in (JAFEE) were used for the leave-one-out experiment.

2) *10-fold cross validation analysis:* We randomly partitioned the dataset into k subsets and then repeated the experiment k times. Each time, the $k-1$ subsets were used for training the classifier while the k^{th} was used for validation. We used different radius R in our analysis. The results show that the accuracy gradually increases as R increases, and the best result is obtained when ($R = 5$), which achieves a high recognition rate of 97.61% as shown in Table I, which also compares the performances of the RBF, linear and polynomial kernels.

TABLE I: Recognition rate of the fold cross validation on the (JAFEE) database using different R.

Radius	RBF-kernel (%)	Linear-kernel (%)l	Polynomial-kernel (%)
2	90.4762	86.6667	95.2381
3	95.2381	90.4762	95.2381
4	95.2381	97.6190	97.6190
5	**97.6190**	**97.6190**	**97.6190**
6	97.6190	95.2381	95.2381

3) *Leave-one-person-out cross validation:* The classifier was trained using $N-1$ users, and evaluated on N user. The results of using different radius R are shown in Table II,

Comparisons with related work are presented in Table III and IV. The statistics of the other methods are taken from [4]. It can be seen that our algorithm outperforms other methods and achieves a classification accuracy of 80.95% using a linear kernel. We used Matlab R2016b running on windows 10 with Intel Core $i7$ CPU at 3.60 GHz. The average computation time to extract the micro- and macro-features for one facial image is 0.097 seconds.

TABLE II: Recognition rate of the leave-one-person-out cross validation on the (JAFEE) database using different R.

Radius	RBF-kernel (%)	Linear-kernel (%)l	Polynomial-kernel (%)
2	54.7619	61.9048	55.6324
3	67.4242	70.4969	67.4242
4	72.0779	71.9697	72.0779
5	**79.6066**	**80.9524**	**77.2257**
6	72.8778	72.87781	71.4286

C. Experiments on CK database

We also tested the different cross validation parameters on the CK database [22]. The database contains images of 100 university students from different countries. There are 593 images with 6 expressions i.e. $\{$anger, disgust, fear, joy, sadness and surprise$\}$. For cross validation, we selected 324 images, which were the last sequences of each subject. 2-, 5-, 7- and 10-fold were employed for evaluation and the recognition rates exceed 98%.

TABLE III: Recognition rate of the 10-fold cross validation on the JAFEE database.

Ref.	Algorithm	Recognition rate %
[25]	BoostLBP	66.8
[9]	Gabor filter	71.9
[25]	LBP	72.4
[25]	BoostLBP	74.2
[26]	BoostLBP	81.0
[36]	LBP	83.25
[34]	LBP	85.71
[1]	SLGS	88.09
[17]	ASM	89.5
[26]	Gabor filter	91.0
[26]	Gabor filter	91.9
[10]	DCT+Gabor+Wavelets+Gasussian Distribution	N/A 93.40
[8]	LBP	93.80
[4]	LPQ+es-LBP-s	96.19
Proposed	micro+macro	**97.61**

TABLE IV: Recognition rate of the leave-one-person-out cross validation on the (JAFEE) database.

Ref.	Algorithm	Recognition rate %
[37]	LBP	53.8
[30]	Gabor histogram	58.7
[11]	LBP	62.9
[3]	CT	63.81
[37]	LBP	65.7
[37]	LBP	65.71
[1]	SLGS	66.66
[3]	ASM	68.5
[31]	LBP+Gabor	70.0
[32]	Gabor filter	72.0
[33]	MB-LGBP	74.18
[4]	LPQ+es-LBP-s	76.67
Proposed	micro+macro	**80.95**

Our experimental results and analysis on both databases demonstrate that the proposed framework enhances the accuracy by considering salient patches, micro-macro patterns to capture key information, which significantly improves accuracy and time performance.

IV. Conclusions and Future Work

We introduce a new framework for facial expression identification. Our approach encodes micro- and macro-patterns from salient patches in facial images. A SVM is used for expression classification. Experimental results using the JAFEE dataset show that our method outperforms existing methods, achieving a recognition rate of 97.61% in 10-fold cross validation and 80.95% on leave-one-person-out cross validation. In future work, we aim to extend our method to expression recognition in videos. We also aim to study dimensionality reduction to further reduce computational time.

References

[1] Abdullah, Mohd Fikri Azli, Md Shohel Sayeed, Kalaiarasi Sonai Muthu, Housam Khalifa Bashier, Afizan Azman, and Siti Zainab Ibrahim. "Face recognition with symmetric local graph structure (slgs)." Expert Systems with Applications 4. 14 (2014): 6131-6137.

[2] Ahonen T, Hadid A, Pietik? caïnen M. "Face recognition with local binary patterns." European conference on computer vision. Springer Berlin Heidelberg, 2004.

[3] Cai, Lin-Bo, and Zi-Lu Ying. "A new approach of facial expression recognition based on Contourlet Transform." 2009 International Conference on Wavelet Analysis and Pattern Recognition. IEEE, 2009.

[4] Chao, Wei-Lun, Jian-Jiun Ding, and Jun-Zuo Liu. "Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection." Signal Processing 117 (2015): 1-10.

[5] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines." ACM Transactions on Intelligent Systems and Technology (TIST) 2.3 (2011): 27.

[6] Chen, Junkai, Zenghai Chen, Zheru Chi, and Hong Fu. "Dynamic texture and geometry features for facial expression recognition in video." In Image Processing (ICIP), 2015 IEEE International Conference on, pp. 4967-4971. IEEE, 2015.

[7] Cossetin, Marcelo J., Julio C. Nievola, and Alessandro L. Koerich. "Facial expression recognition using a pairwise feature selection and classification approach." Neural Networks (IJCNN), 2016 International Joint Conference on. IEEE, 2016.

[8] Feng, Xiaoyi, M. Pietikainen, and Abdenour Hadid. "Facial expression recognition with local binary patterns and linear programming." Pattern Recognition And Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii 15.2 (2005): 546.

[9] G. Guo, R. Dyer, Simultaneous feature selection and classifier training via linear programming a case study for face expression recognition, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2003, pp. I-346???I-352.

[10] Gupta, Sandeep K., et al. "A hybrid method of feature extraction for facial expression recognition." Signal-Image Technology and Internet-Based Systems (SITIS), 2011 Seventh International Conference on. IEEE, 2011.

[11] Huang, Ming-Wei, Zhe-wei Wang, and Zi-Lu Ying. "A new method for facial expression recognition based on sparse representation plus LBP." Image and Signal Processing (CISP), 2010 3rd International Congress on. Vol. 4. IEEE, 2010.

[12] Jain, Anil K., and Stan Z. Li. Handbook of face recognition. New York: springer, 2011.

[13] Jing, Xiao-Yuan, Hau-San Wong, and David Zhang. "Face recognition based on 2D Fisherface approach." Pattern Recognition 39.4 (2006): 707-710.

[14] Jolliffe, Ian. Principal component analysis. John Wiley and Sons, Ltd, 2002.

[15] Kanade, Takeo, Jeffrey F. Cohn, and Yingli Tian. "Comprehensive database for facial expression analysis." Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on. IEEE, 2000.

[16] Kotsia, Irene, Stefanos Zafeiriou, and Ioannis Pitas. "Texture and shape information fusion for facial expression and facial action unit recognition." Pattern Recognition 41.3 (2008): 833-851.

[17] Lei, Gang, et al. "Geometric feature based facial expression recognition using multiclass support vector machines." Granular Computing, 2009, GRC'09. IEEE International Conference on. IEEE, 2009.

[18] Liao S, Fan W, Chung AC, Yeung DY. "Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features." 2006 International Conference on Image Processing. IEEE, 2006.

[19] Lu, Jiwen, and Yap-Peng Tan. "Regularized locality preserving projections and its extensions for face recognition." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 40.3 (2010): 958-963.

[20] Lyons, Michael J., Julien Budynek, and Shigeru Akamatsu. "Automatic classification of single facial images." IEEE Transactions on Pattern Analysis and Machine Intelligence 21.12 (1999): 1357-1362.

[21] Pantic M and Patras . "Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 36.2 (2006): 433-449.

[22] Ojansivu, Ville, and Janne Heikkil?¢ã. "Blur insensitive texture classification using local phase quantization." International conference on image and signal processing. Springer Berlin Heidelberg, 2008.

[23] Ojala T, Pietik?¢ãinen M, Harwood D. "A comparative study of texture measures with classification based on featured distributions." Pattern recognition 29.1 (1996): 51-59.

[24] Seshadri, K., and M. Savvides. "Towards a Unified Framework for Pose, Expression, and Occlusion Tolerant Automatic Facial Alignment." IEEE transactions on pattern analysis and machine intelligence (2015).

[25] Shan, C., Gong, S., and McOwan, P. W. (2006, June). A comprehensive empirical study on linear subspace methods for facial expression analysis. In 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06) (pp. 153-153). IEEE.

[26] Shan, Caifeng, Shaogang Gong, and Peter W. McOwan. "Facial expression recognition based on local binary patterns: A comprehensive study." Image and Vision Computing 27.6 (2009): 803-816.

[27] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.

[28] Walecki R, Rudovic O, Pavlovic V, Pantic M. "Variable-state Latent Conditional Random Field models for facial expression analysis." Image and Vision Computing (2016).

[29] Wu T, Bartlett MS, Movellan JR. "Facial expression recognition using gabor motion energy filters." 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. IEEE, 2010.

[30] Xiaomin, Liu, and Zhang Yujin. "Facial expression recognition based on Gabor histogram feature and MVBoost." Journal of Computer Research and Development 7 2007.

[31] Zavaschi, Thiago HH, Alessandro L. Koerich, and L. E. S. Oliveira. "Facial expression recognition using ensemble of classifiers." 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2011.

[32] Zhang, Zheng, Chi Fang, and Xiaoqing Ding. "Facial expression analysis across databases." Multimedia Technology (ICMT), 2011 International Conference on. IEEE, 2011.

[33] Zhang, Zheng, Chao Xu, Jiaxin Wang, and Xiangning Chen. "Facial Expression Recognition Based on MB-LGBP Feature and Multi-level Classification." Advances in Multimedia, Software Engineering and Computing Vol. 2. Springer Berlin Heidelberg, 2011. 37-42.

[34] Zhang, Shiqing, Xiaoming Zhao, and Bicheng Lei. "Facial expression recognition using local fisher discriminant analysis." International Conference on Computer Science, Environment, Ecoinformatics, and Education. Springer Berlin Heidelberg, 2011.

[35] Zhang X and Mahoor MH. "Task-dependent multi-task multiple kernel learning for facial action unit detection." Pattern Recognition 51 (2016): 187-196.

[36] Zhao, Xiaoming, and Shiqing Zhang. "Facial Expression Recognition via Fuzzy Support Vector Machines." Advances in Computer Science and Information Engineering. Springer Berlin Heidelberg, 2012. 495-500.

[37] Zilu, Ying, and Fang Xieyan. "Combining LBP and Adaboost for facial expression recognition." 2008 9th International Conference on Signal Processing. IEEE, 2008.