Classification of Mouth Action Units using Local Binary Patterns

Sarah Adel Bargal
The American University in Cairo
Department of Computer Science, AUC, AUC Avenue, P.O. Box 74 New Cairo 11835, Egypt
s_bargal@aucegypt.edu

Rana el Kaliouby
Massachusetts Institute of Technology
Media Laboratory, MIT, 20 Ames Street, Cambridge USA
kaliouby@media.mit.edu

Amr Goneid
The American University in Cairo
Department of Computer Science, AUC, AUC Avenue, P.O. Box 74 New Cairo 11835, Egypt
goneid@aucegypt.edu

Anas Nayfeh
Kuwait University
Center of Information Systems, KU, Shuwaikh, Kuwait
anas.nayfeh@ku.edu.kw

Abstract

Automated facial expression recognition remains a grand challenge in computer vision. This paper lays the groundwork for building a system that automatically, and in real-time, accurately classifies Action Units (AUs) of the Facial Action Coding System. We use Local Binary Patterns (LBPs) for the feature extraction stage, followed by support vector machines for classification. LBPs require no manual initializations, are computationally simple, run in real-time, are illumination tolerant, and do not require high image resolutions. Experiments compare the performance of LBP to that of image quantization in the spatial and frequency domains. Results show that using LBPs for feature vector extraction of mouth AUs yielded the highest average classification rate of 80%, suggesting that such LBP feature vectors may be used for classifying more AUs and AU combinations. Results obtained are consistent for template matching and support vector machines.

1. Introduction

Automated facial expression classification has many applications including Human-Computer Interaction, psychological and computer vision research, and medicine. Examples include interactive computer games, smarter interfaces, automatic database labeling, automatic monitoring of patients' moods, and driver state detection. Most of these applications require real-time implementations.

Ekman and Friesen's Facial Action Coding System (FACS) [5] is a comprehensive coding system for describing facial expressions. Muscle movements are coded as action units (AUs) and facial expressions are coded using one or more AUs. Humans produce thousands of expressions, all of which can be objectively described using FACS. Thus, FACS has been used extensively for the automated analysis of facial expressions, e.g., [3,6,15,19], and is more descriptive than systems that only recognize a limited set of basic emotions.

Automated facial expression recognition consists of three main stages after image pre-processing: extracting the feature vector which describes the facial image, reducing the dimensionality of that vector if possible, and finally, classifying test images into certain classes after training using the reduced dimensionality vector. Feature extraction is key, because finding a descriptive, representative, efficient, and easy to compute feature vector of the facial image will largely impact the classification results.

This paper describes feature extraction and classification of mouth AUs from static images. Seyedarabi *et al.* [21] emphasize that the mouth has the most flexible deformability, while Liu and Wang [16] assert that the mouth contributes the most to facial expressions. The principal contribution of the paper is the feature extraction of mouth AUs using Local Binary Patterns (LBPs), a feature extraction technique that is efficient, is automated and runs in real-time, while maintaining classification accuracy. LBPs are also illumination tolerant, and do not require high image resolutions and are therefore suited to real-world contexts and environments.

The paper is organized as follows: section 2 surveys feature extraction methods for face analysis, as well as related work that use LBPs; section 3 gives an overview of Local Binary Patterns for representing mouth AUs. Section 4 describes classification using template matching and support vector machines; section 5 present experimental evaluation and result, while section 6 concludes the paper.

2. Related work

Feature vector extraction for AU classification has been addressed using several methodologies (see [7,20] for a survey). More recently, Seyedarabi *et al.* [21] classify mouth AUs by using a two-step active contour for feature extraction. Active contours do not run in real-time and require many parameter initializations. Pantic and Patras [19] use feature point tracking to detect 27 AUs, which mainly considers lined facial features. El Kaliouby and Robinson [12] perform recognition of affective and cognitive states using feature point tracking. Lien *et al.* [15] use feature point tracking, dense flow tracking, and gradient component analysis to detect upper-face AUs. Gradient component analysis mainly considers furrows and is best suited for areas like the forehead and cheeks. Dense flow tracking is not real-time. Fasel and Luettin [6] describe a system that recognizes AUs by obtaining difference images and projecting them once in the Principal Component Analysis (PCA) sub-space, and another in the Independent Component Analysis (ICA) subspace. Whitehill and Omlin [23] use Haar features and Gabor filters to detect AUs. Although Gabor wavelets produce higher recognition rates, they are both slow and memory inefficient.

LBP is a technique that is used for texture description [18]. Since 2004, LBP has been applied successfully to face detection and recognition [1, 2, 11], where it outperformed methods like PCA. Lian *et al.* [14] use LBP for gender classification with very high recognition rates. Emotion recognition (for the six basic emotions) using LBP is addressed holistically (on the whole face) giving impressive results [8–10]. Because LBPs are a simple computational operator that is real-time, and produces accurate results when applied to face recognition, gender classification, and emotion recognition, this has motivated us to apply LBP to facial AU classification.

3. Local Binary Patterns

LBPs have recently gained attention, outperforming other methods because of its performance and computational efficiency. Also, LBP is illumination tolerant [18], robust to parameter selection in terms of performance [1], does not require initial points, and performs reliably over a range of low image resolutions [22].

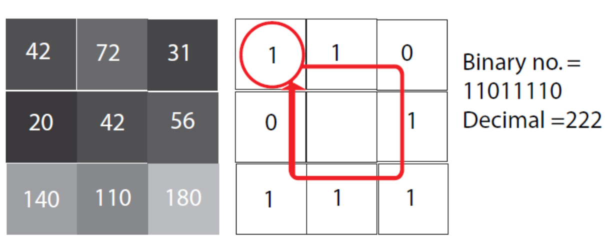

Figure 1. LBP codes are computed by thresholding on a central pixel and taking the decimal equivalent of the binary number.

Initially, a neighborhood of a certain size is defined. The original LBP neighborhood [17] defined is 3*3. The central pixel is used to threshold surrounding pixels, producing an LBP Code (Fig 1). This was then extended to a neighborhood of P points for radius R; the notation *LBP*(*P;R*) is used. Each pixel is replaced by the decimal equivalent (LBP Code) of a P-bit binary pattern. The LBP histogram of the image depicts the frequencies of all possible LBP Codes. The histogram frequencies are represented as in Eq. 1. $I(x, y)$ represents the image pixels at co-ordinates x and y, n represents the number of LBP Codes (bins: uniform or non-uniform), and $F\{A\} = 1$ if A is true, and is zero otherwise.

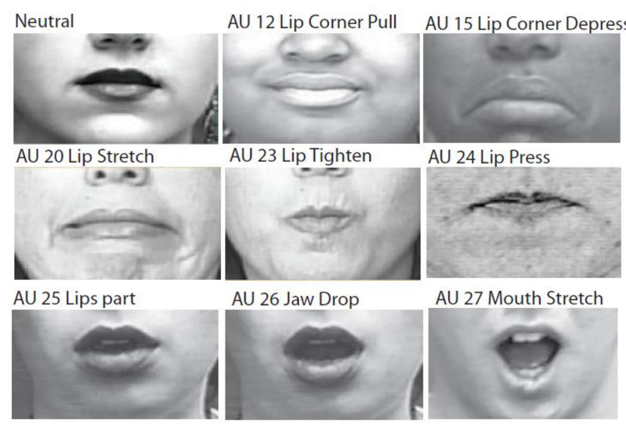

Figure 2. Mouth AUs that are addressed in this paper (examples are from the Cohn-Kanada database [13].

$$H_i = \sum_{x,y} F\{I(x,y) = i\}, i = 0, \ldots, n-1 \qquad (1)$$

As an extension, uniform patterns are introduced by Ojala *et al.* [18], reducing the dimensionality of the vector representing the histogram. A U2 uniform pattern is a sequence of zeros and ones that contains no more than two zero-to-one or one-to-zero transitions. The notation $LBP^{u2}_{(P,R)}$ is used for LBP using U2 uniform patterns. Ojala *et al.* [18] noticed that about 90% of patterns in texture images are uniform using the standard (8,1) neighborhood, and that the percentage decreases as the neighborhood grows. For further encoding, the image is segmented into regions. The feature vector would be the concatenation of all LBP histograms of all regions. This method produces a longer feature vector, however, it is more descriptive of the image; a trade-off exists. Eq. 2 shows the notation used for the concatenated histograms representing the feature vector for m regions, and $i = 0, \ldots, n-1$ and $j = 0, \ldots, m-1$:

$$H_{i,j} = \sum_{x,y} F\{I(x,y)=i\}F\{I(x,y)\epsilon R_j\} \qquad (2)$$

4. AU classification

The mouth AUs listed in Fig. 2 are among the most frequently occurring lower-face AUs and are the ones we address in this paper: AU12 is a lip corner pull; AU15 is a lip corner depress; AU20 is a lip stretch; AU23 is a a lip tighten and AU24 represents a lip press; AU25 is a lips part; AU26 is a jaw drop while AU27 is a mouth drop.

4.1. Pre-processing

Faces are detected in training and testing images, and scale normalization is conducted. After that, the face is cropped twice such that only the mouth region remains. This is done after determining two crop fractions statistically from 50 subjects.

Table 1. Eight basic feature vectors used for experimentation. In all eight cases a neighborhood of P = 8 was used. The last column gives the feature size per region.

#	Image	#regions	Mapping	Eq.	Size/region
1	Raw	1	none	eq.1	256
2	Raw	1	u2	eq.1	59
3	Raw	9,36	none	eq.2	256
4	Raw	9,36	u2	eq.2	59
5	Difference	1	none	eq.1	256
6	Difference	1	u2	eq.1	59
7	Difference	9,36	none	eq.2	256
8	Difference	9,36	u2	eq.2	59

4.2. Feature Vector Extraction and Dimensionality Reduction

LBP is used for feature vector extraction from a mouth image or a difference image. The difference image is the difference between an image of a subject possessing a certain AU and the subject's neutral image. We experiment with both (8,1) and (8,2) LBP neighborhoods. We also experiment using LBP on the whole mouth/difference image, LBP on the image divided into 9 regions, and LBP on the image divided into 36 regions. Moreover, experiments are conducted both with and without using uniform patterns for dimensionality reduction. The eight basic LBP vectors are summarized in Table 1.

Fig. 3 shows a graphical representation of an LBP feature vector ($LBP^{u2}_{(P,R)}$, 9 regions) on the average template of every class. Note that AU23 and AU14 have been merged into one class, since they co-occur in about 80% of the images that contain either AUs. The graphical representation shows a unique signature for every class template. As expected, AUs that are similar in appearance, e.g., AU12 and AU15 have closer signatures.

LBP is compared against the basic feature extraction technique, image quantization. To ensure fair comparison, we performed many experiments (Fig. 4) using various parameter settings for both LBP and image quantization in order to find the best performance of each. As illustrated in Fig. 4, image quantization is performed using both mouth images, and difference mouth images. Also, quantization is performed in both the spatial and frequency domains, obtained by once using fast fourier transform (FFT), and another using discrete cosine transform (DCT)). Moreover, six different quantization window sizes are used; 15*15, 20*20, 25*25, 30*30, 35*35, 40*40. The mean of a window is used to represent it.

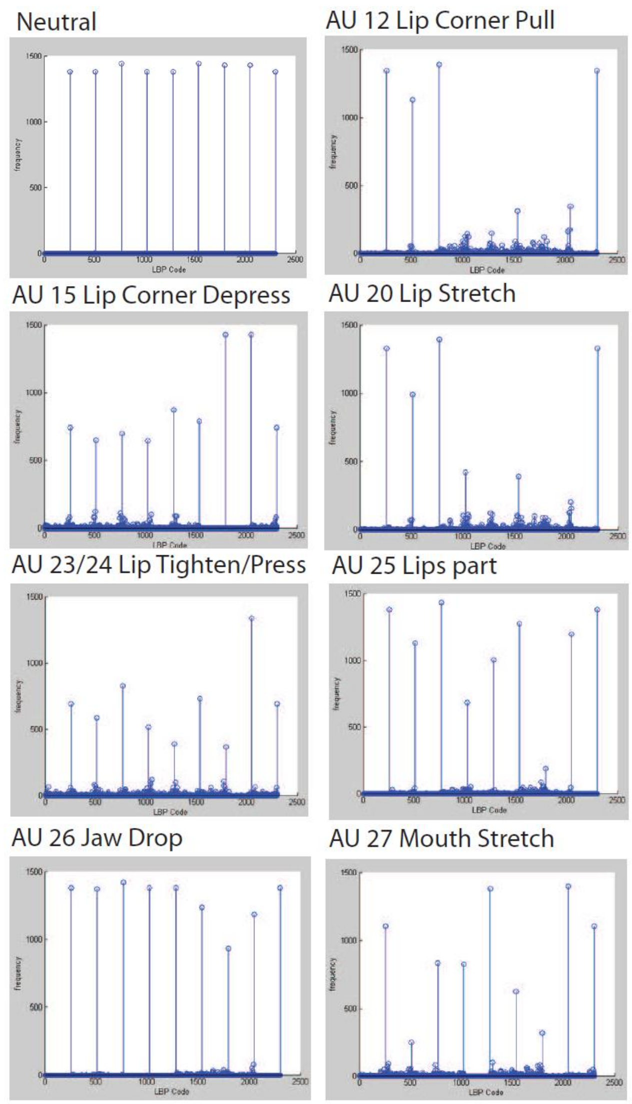

Figure 3. LBP features ($LBP^{u2}_{(P,R)}$, 9 regions using differences images) for mouth AUs.

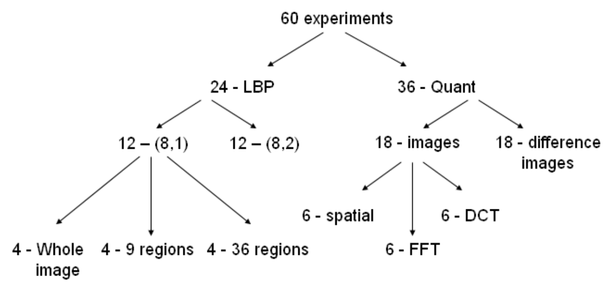

Figure 4. LBP and quantization experiments for 7-class and 8-class classifications on raw and difference images.

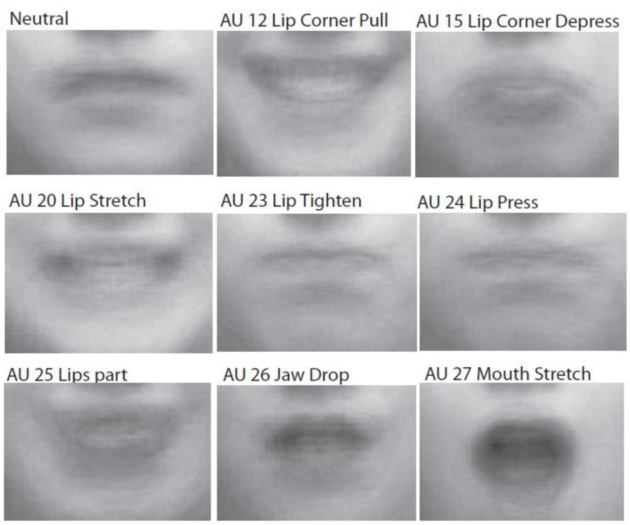

Figure 5. Templates of mouth AUs.

4.3. Classification

Two classification methods are used; template matching and support vector machines (SVMs). SVMs embed data into high dimensional feature spaces such that they are separable using simple linear algebra and geometry rules. A template for every AU is obtained by averaging the feature vectors of images used for training. Direct averaging is used for template construction. A graphical representation of AU templates is obtained by superimposing images of every AU, shown in Fig. 5. For classification, a nearest neighbor classifier using Euclidean distance is used. We use the LIBSVM Matlab library [4]to train an SVM classifier per mouth AU that takes as input the LBP feature vectors and their corresponding labels. The trained SVM then is used to classify unseen LBP feature vectors.

Classification is done once using the 7 AU classes in Fig. 5 and another by adding the eighth neutral class. The neutral class is added to show the ability of encoding subtle mouth movements. The neutral mouth can be easily confused with lip tight or lip press if subtle details are not encoded, affecting recognition percentages.

5. Evaluation

AU classifiers should be robust to scale (size of the image), and to age, ethnicity, and gender of the subjects. Scale tolerance is accounted for at early pre-processing stages. Age, ethnicity, and gender tolerance are dependent on the database used. A database that is well-suited to the purpose of this work is the Cohn-Kanade database [13]. Fig. 4 depicts the 60 basic experiments undergone to investigate various parameters and applications of both LBP and quantization. These 60 experiments were performed once using 7-class classification, and another using 8-class classification; total of 120 experiments. 24 out of the 60 experiments are LBP experiments. In 12 of the 24 experiments an (8,1) neighborhood is used, an (8,2) neighborhood is used in the rest. Every 12 experiments are divided into 3 groups of 4. The first group is undergone using feature vectors 1,2,5,6 (table 1) on the whole image, the second using feature vectors 3,4,7,8 on the image divided into 9 regions, and the third using feature vectors 3,4,7,8 on the image divided into 36 regions. The best performing feature vectors are then classified using SVMs to make sure that the method is classifier independent.

5.1. Implementation Details

A k-fold (k=5) cross-validation and subject independency are used for testing. The performance measure used is the recognition rate resulting from the classification procedure. A confusion matrix is produced; the average of 5 matrices of 5 folds. For every AU in this confusion matrix, the detected, undetected (false negative) and falsely detected (false positive) percentages are calculated.

5.2. Experimental Results

The detected, undetected, and false positive rates are obtained for every experiment. Using difference images in both LBP and quantization experiments yielded consistently higher (4-15%) recognition rates than using the AU image. Regarding LBP, dividing the image into regions clearly yields more accurate descriptions than applying LBP to the whole image. Dividing the image into 9 regions and performing LBP on each shows higher recognition rates than dividing the image into 36 regions. Unnecessary divisions cause details to disappear from the regions. Using U2 mapping gave recognition rates that are very similar to (+/-1%) the unreduced vector. The reduced U2 feature vector however is shorter, and therefore would be classified better. The (8,1) neighborhood gave recognition rates 3% higher on average than the (8,2) neighborhood. The optimal LBP feature vector obtained by the presented experiments is the one that divides difference images into 9 regions, performing LBP on every region using uniform patterns and an (8,1) neighborhood. The detected, undetected, and false positive rates of the best LBP feature vector is depicted in Fig. 6.

Fig. 7 shows the number of labeled images containing various mouth AUs in the Cohn-Kanade Database. The distribution in Figs. 6 and 7 appear to have the same shape. We note that the classes with high recognition rates are the ones with a high number of images available for training and testing. The distribution is the same regarding quantization. Therefore, classes will be divided into two groups; a group with a sufficient number of images (> 55) and another with a limited number of images (AU23/24 and AU26). Results are shown in Table 2. All false positive rates range from 2.3% to 10.5%, LBP having an acceptable rate of 5.5%.

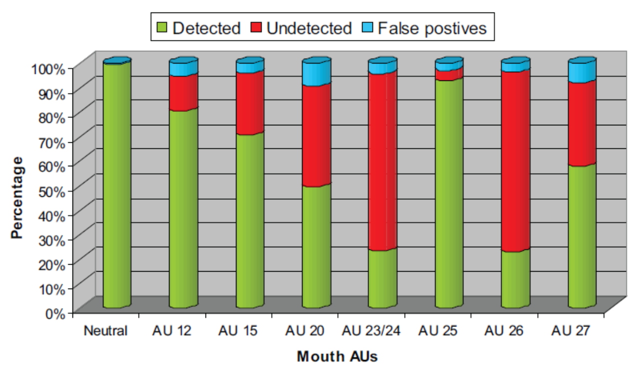

Figure 6. Average rate for correctly detected, undetected, and false positives for each feature extraction approach.

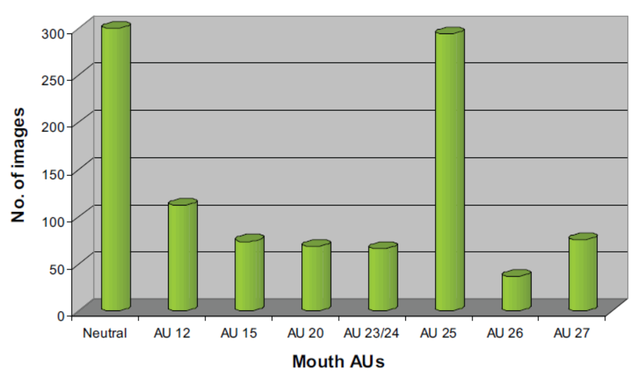

Figure 7. The number of labeled images available in the Cohn-Kanade Database for each of the mouth AUs.

Table 2 shows the detection and false positive rates of the best performing spatial, FFT, DCT, and LBP feature vectors. The percentages shown are the average of both 7-class and 8-class classification. Regarding classes with sufficient numbers of images, LBP showed the highest average detection rate of 80%. We note that LBP detection rate of 8-class classification is 82%, whereas that of the 7-class classification is 78%; giving an average of 80%. This indicates that the addition of a class did not confuse the classifier further. Also, using the LBP feature vector that gave the highest detection percentages on a simple design SVM classifier, a similar 78% detection rate is obtained. This indicates that using LBP is not classifier dependent. Using a more sophisticated SVM classifier design is expected to further raise detection rates.

Table 2. Detected or True Positive (TP) and False Positive (FP) percentages obtained for classes with sufficient and limited numbers of training images.

Classification	Sufficient		Limited	
	TP	FP	TP	FP
LBP	80	7	33	5.5
Spatial	70	8	22	6.5
DCT	60	9	24	2.3
FFT	57	1.5	33	10.5

6. Conclusion

Facial expression classification remains a challenging task for machines. To the best of our knowledge, experimentation with LBP on specific facial features, and not on the whole face, with the aim to recognize single AUs, and not emotions, is a novel approach. The method imple mented in this work is advantageous over existing feature extraction methods that are slow, require manual initialization, are illumination dependent, or require high image resolution. LBP features are extracted quickly, however they result in an average recognition percentage of 80% for classes with sufficient image numbers. LBP showed a higher detection percentage than image quantization in both the spatial and frequency domains. The percentage is also robust to two classification methods, and is person independent.

Future work includes extending our work to support more AUs and AU combinations using LBP for facial feature extraction, as well as incorporating temporal information. We also plan to test the robustness of LBPs to pose changes and head motion, which often co-occur with facial expressions, as well as its performance with spontaneous versus posed expressions and its generalization across datasets.

ACKNOWLEDGEMENTS

We would like to thank Professors J.F. Cohn and T. Kanade for providing us with the Cohn-Kanade Facial Expression Database.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. *European Conference on Computer Vision (ECCV)*, 2004.

[2] T. Ahonen, A. Hadid, and M. Pietikainen. Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[3] M. Bartlett, L. G., I. Fasel, and J. Movellan. Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction. In *CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction*, 2003.

[4] C. Chang and C. Lin. LIBSVM: a library for support vector machines. *Software available at http://www. csie. ntu. edu. tw/cjlin/libsvm*, 80:604–611, 2001.

[5] P. Ekman and W. Friesen. *Facial Action Coding System*.

[6] B. Fasel and J. Luettin. Recognition of Asymmetric Facial Action Unit Activities and Intensities. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, volume 1, pages 1100–1103, 2000.

[7] B. Fasel and J. Luettin. Automatic Facial Expression Analysis: A Survey. *Pattern Recognition*, 36:259–275, 2003.

[8] X. Feng. Facial expression recognition based on local binary patterns and coarse-to-fine classification. *Computer and Information Technology, 2004. CIT'04. The Fourth International Conference on*, pages 178–183, 2004.

[9] X. Feng, A. Hadid, and M. Pietik¨ainen. A Coarse-to-Fine Classification Scheme for Facial Expression Recognition. *ICIAR Proceedings, Lecture Notes in Computer Science*, 3213:668–675.

[10] X. Feng, M. Pietik¨ainen, and A. Hadid. Facial Expression Recognition with Local Binary Patterns and Linear Programming. *Pattern Recognition and Image Analysis*, 15(2):546–548, 2005.

[11] P. M. Hadid, A. and T. Ahonen. A discriminative feature space for detecting and recognizing faces. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[12] R. el Kaliouby and P. Robinson. *Real-Time Vision for Human Computer Interaction*, chapter Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures, pages 181–200. Springer-Verlag, 2005.

[13] T. Kanade, J. Cohn, and Y.-L. Tian. Comprehensive Database for Facial Expression Analysis. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, pages 46–53, 2000.

[14] H. Lian and B. Lu. Multi-view Gender Classification Using Local Binary Patterns and Support Vector Machines. *Proceeding of the 3rd International Symposium on Neural Networks ISNN06*, 2:202–209.

[15] J. J. Lien, A. Zlochower, J. F. Cohn, and T. Kanade. Automated Facial Expression Recognition Based on FACS Action Units. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, pages 390–395, 1998.

[16] W. Liu and Z. Wang. Facial expression recognition based on fusion of multiple gabor features. In *International Conference on Pattern Recognition*, 2006.

[17] T. Ojala, M. Pietikainen, and D. Harwood. Comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[18] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[19] M. Pantic and I. Patras. Detecting Facial Actions and their Temporal Segments in Nearly Frontal-View Face Image Sequences. *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, 4, 2005.

[20] M. Pantic and L. Rothkrantz. Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.

[21] H. Seyedarabi, W. Lee, and A. Aghagolzadeh. Automatic lip tracking and action units classification using two-step active contours and probabilistic neural networks. In *Canadian Conference on Electrical and Computer Engineering (CCECE/CCGEI'06)*, 2006.

[22] C. Shan, S. Gong, and P. McOwan. Robust facial expression recognition using local binary patterns. In *IEEE International Conference on Image Processing*, 2005.

[23] J. Whitehill and C. Omlin. Haar Features for FACS AU Recognition. *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06)- Volume 00*, pages 97–101, 2006.