Person-independent Facial Expression Recognition via Hierarchical Classification

Mingliang Xue [1], Wanquan Liu [2], Ling Li [3]

Department of Computing, Curtin University
Perth WA 6102, Australia
[1] mingliang.xue@postgrad.curtin.edu.au
[2] W.Liu@curtin.edu.au
[3] L.Li@curtin.edu.au

***Abstract*—Automatically recognizing facial expressions presents an active and challenging problem in computer vision and pattern classification. The person-independent case is even more challenging. In this paper, we propose a hierarchical approach to achieve person-independent facial expression recognition. Specifically, the expressions that are easily confused together are merged into one class and join the remaining prototypic expressions in the first tier classification; the expressions in the merged class are then separated in the second tier. Support Vector Machine is adopted as the classifier in both tiers, with the LBP and displacement features in the first tier as well as mouth and eyebrows features in the second tier. The proposed metLhod is tested on the Cohn-Kanade Extended (CK+) dataset and evaluated in terms of a confusion matrix. The person-independent experiments demonstrate the effectiveness of the proposed hierarchical classifier in improving recognition accuracy and eliminating confusions.**

I. Introduction

Facial expressions provide one of the most powerful and natural means for human beings to communicate their emotions and intensions. The possibility of enabling computers to recognize and analyze the information conveyed by facial expression has attracted significant research interest over the last few years. It has given rise to a number of methods for automatic facial expression recognition [1][2]. Though much progress has been made, robust and effective facial expression recognition remains difficult due to the subtlety and variability of facial expressions. These studies conducted either person-independent or person-dependent facial expression recognition in the experiments. The person-independent recognition means the individuals included in the testing images never show up in the training images, while in the person-dependent recognition the individuals appeared in the testing images also show up in the training images. Obviously, a feasible facial expression recognition system should be able to work person-independently. The comparisons of the two cases in [3] reveal that the person-independent case is much more difficult than the person-dependent case. So considerable researches, as well as this paper, focus on the person-independent facial expression recognition.

The most commonly used study on classification of facial expression is Ekman's cross-cultural study [4] on the existence of *universal categories of emotional expressions*, which is referred to as the prototypic expressions consisting of happiness, sadness, surprise, fear, anger and disgust. It has been shown that the six prototypic expressions are not mutually distinguishable. The systems in [3][5][6][7] show that there are often confusions between anger and disgust while in [8][9] sadness is often confused with anger. In order to achieve highly accurate recognitions for all the expressions, many researchers attempted to eliminate such confusions. The attempts generally involve two vital steps: feature extraction and expression classification.

There are two common approaches to extract facial features from the original face images: geometric feature-based approach [10][11] and appearance-based approach [12][13]. In[14][15], the authors argued that appearance features are better than geometric features, because geometric features are more sensitive to inaccurate image alignment. In addition, Lucey et al. [16] showed that appearance information is more important to the recognition of anger, sadness and fear. However, with the recent development on face alignment and facial feature tracking, an increasing number of expression analysis algorithms are based on geometric features. Valstar et al. [11] presented a method that can detect facial action units effectively by classifying features derived from the tracked facial landmarks. They argued that the geometric features is well suited for facial expression analysis, especially with facial feature tracking. The studies from both sides indicate a possible combination of these two kinds of features as a better face representation for facial expression recognition.

Normally, a classifier is designed based on the extracted features for the classification of the six prototypic expressions, in which all the expressions are treated equally and evenly. Out of the six expressions, happiness and surprise are the easiest to recognize [7][17]. The remaining four expressions are more subtle ones that are often confused with each other. This fact prompts us to divide and conquer the problem of recognizing the 6-classes expressions by a hierarchical classification. The expressions that are commonly confused are merged into one class, which is joined by the remaining prototypic expression classes to form the first tier of classification. It is expected that the first tier classification will perform well since the expressions that easily confused have been merged together.

In the second tier, the prototypic expressions in the merged class are separated by an earmarked classifier. The hierarchical classification provides an opportunity to utilize different features to obtain the best performance in each tier.

In this paper, a hierarchical SVM classifier is designed to improve the performance of person-independent facial expression recognition. This hierarchical classifier enables us to divide and conquer the recognition problem in two tiers. The easily-confused expressions are merged into one class in the first tier and then separated in the second tier. We also propose to utilize different kinds of features in each tier, because classification of different expressions is targeted in each tier. The combined features of LBP and displacement are used for facial expression description in the first tier since it yields the best performance among possible features. In the second tier, the landmarks on mouth and eyebrows are selected to represent expressions since mouth and eyebrows are proven highly related to those easily-confused expressions[18][6]. The experimental results obtained by applying the hierarchical classifier on the CK+ dataset [16] demonstrate the satisfactory performance of the proposed method.

The remainder of this paper is organized as follows. Section II discusses some background knowledge for facial expression recognition. In section III the details of the proposed hierarchical classification method are presented. We record the experimental results of the hierarchy classification in section IV and conclude in section V.

II. Background information

A. Facial expression representation

Effective facial feature extraction from face images plays an important role in facial expression recognition. The appearance features and the geometric features are the common features utilized in facial expression representation.

Appearance feature: The appearance features model the appearance changes of faces, mainly caused by different facial expressions. Shan et al. [13] compared the LBP features with the Gabor features for facial expression recognition using different classifiers, and studied their performances over various resolutions. The results revealed that the LBP features are effective and efficient for facial expression recognition, even for low resolution face images.

In this paper, we adopt the LBP for facial appearance feature representation. Accumulating the uniform patterns [13] yields an LBP operator $LBP^{u2}_{P,R}$, where the superscript $u2$ stands for using only uniform patterns in P equidistant sampling points on a circle of radius of R. The $LBP^{u2}_{P,R}$ has fewer labels than the original $LBP_{P,R}$ operator. For example, the number of labels in the neighbourhood of $(8, 1)$ is 256 for the original $LBP_{8,1}$ but 59 for $LBP^{u2}_{8,1}$, which is a big relief on computational complexity.

The LBP histogram contains appearance information about the distribution of the local micro-patterns in facial expression images. Therefore, it can serve intuitively as the appearance feature representation in facial expression analysis.

Geometric features: Geometric features have been widely exploited in facial representation [19][20], where locations and displacements of facial components are extracted to represent the face geometry. Recently, Lucey et al. [16] manually label some keyframes in video sequences and use a descent AAM fitting algorithm [21] to get the landmarks of the remaining frames. Their work shows that facial expression recognition benefits a lot from fusion of both shape and appearance features.

Assuming that the facial components have been labelled with N landmarks, the coordinates of the landmarks could be denoted as $p_i = (x_i, y_i), i = 1, ..., N$. The face images could then be represented by the location information vector $P = [p_1\ p_2\ ...\ p_i\ ...\ p_N]$, which is the concatenation of all the landmarks p_i. The location information vectors encapsulate the shape and position of the facial components which are affected by the expressions. Furthermore, in facial expression recognition based on image sequences, the facial movements can be measured by the geometrical displacement of corresponding facial feature points between the current frame and the initial frame. The displacement vector D can hence be derived from concatenating the displacements of all the landmarks:

$$D = [\Delta x_1\ \Delta y_1\ \Delta x_2\ \Delta y_2\ ...\ \Delta x_i\ \Delta y_i\ \ ...\ \Delta x_N\ \Delta y_N] \quad (1)$$

where $\Delta x_j, \Delta y_j$ is the x, y coordinate displacement of the j-th landmarks respectively. The displacement information encodes the motion of the landmarks from a neutral face to faces with expressions of different intensity. Both the displacement and location information are utilized in our method because they are directly related to facial expressions.

B. Support Vector Machines

Support vector machines (SVMs) have been proven powerful in facial expression classification [11][12]. It also achieves the best performance according to a comprehensive study [13], so we adopt SVMs as the classifiers for facial expression recognition in this paper. SVMs attempt to find the hyperplane that maximizes the margin between the positive and negative observations for a specified class. Given a training set of labelled examples $\{(x_i, y_i), i = 1, ..., k\}$ where $y_i \in \{-1, 1\}$, a testing example x is labelled by the following function:

$$f(x) = sgn(\sum_{i=1}^{k} \alpha_i y_i K(x_i, x) + b) \quad (2)$$

where α_i are Lagrange multipliers of a dual optimization problem that determine the classification hyperplane, $K(\cdot, \cdot)$ is a kernel function, and b is the threshold parameter of the hyperplane.

SVMs make binary decisions. However, there are six classes in facial expression recognition, each representing one of the prototypic expressions (anger, disgust, fear, sadness, happiness and surprise). In this paper, we use LIBSVM [22] for the training and testing of SVMs, which achieves the multi-classes classification according to the one-against-rest technique. With

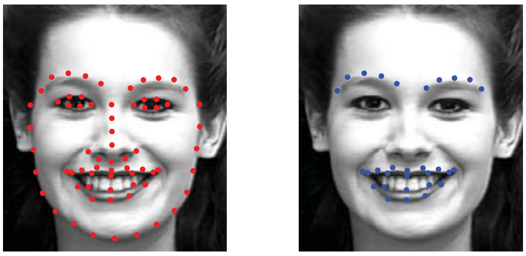

Fig. 1. (Left)Landmarks used in displacement feature extraction. (Right)The selected landmarks of Mouth and Eyebrow.

regard to the parameter selection, we carry out coarse-to-fine grid search in a 5-fold cross-validation. The parameter which yields the best cross-validation accuracy is selected for the decision function.

III. The proposed method

As mentioned before, the six prototypic expressions are not mutually distinguishable, so the confusions caused by the easily-confused expressions will affect the recognition performance significantly. In order to improve the recognition performance, the proposed method attempts to eliminate such confusions via a hierarchical classification. It has two advantages. Firstly, the hierarchical classification can pick the distinguishable expressions out in the first tier, and then focuses on the classification of easily-confused ones in the second tier. Secondly, the hierarchical structure enables us to utilize the most appropriate features for expression recognitions in each tier. This section provides the details of the proposed hierarchical classification method.

A. Feature extraction

In the proposed method, both the appearance-based feature (LBP features) and the geometric feature (Displacement and Location features) are extracted from the face images.

LBP features: Each aligned face image is divided into 42(6×7) blocks, and the 59-bin $LBP_{8,1}^{u2}$ operator is run on each block, which is a trade-off between the recognition performance and computational complexity. The 59-bin LBP histogram derived from each of the 42 blocks are concatenated to a 2478(59×42) dimension vector to represent a face image.

Displacement features: In the proposed method, 68 landmarks which are tracked by Lucey et al.[16] are utilized to extract the displacement information, as illustrated in Fig. 1 (Left). For each face image, the x, y coordinates displacement of the landmarks are obtained by subtracting the landmark locations of the neutral image from the corresponding landmarks locations in current image. The displacement feature is represented by a vector with the length of 136 which is formed by concatenating the x, y coordinate displacements.

MEb(Mouth and Eyebrow) features: The mouth and eyebrows are the most important parts for facial expression recognition. It has been shown by Pardas and Bonafonte [17] that the mouth and eyebrows possess the maximum amount of information related to facial expressions, with the mouth carrying more information than the eyebrows. In [6], Kotsia et al. show that occlusion of the mouth leads to inaccuracies in the recognition of anger, fear, happiness and sadness, whereas the occlusion of the eyes and brows leads to a dip in the recognition accuracy of disgust and surprise. Another occlusion research by Bourel et al. [18] has demonstrated that sadness is mainly conveyed by the mouth. Thus, we select the landmarks on the mouth and eyebrows (see Fig.1 (Right)) and utilize both the location and displacement of these landmarks to form the MEb (Mouth and Eyebrow) feature for distinguishing anger and sadness confusion which are most commonly confused.

B. Hierarchical classifier design

Since the six prototypic expressions are not evenly distinguishable, we attempt to divide and conquer the recognition problem by a hierarchical classification. The hierarchical classification has a structure of two tiers. In the first tier, the easily-confused prototypic expressions are considered as one class and join the remaining expressions for classification. In the second tier, another classifier, which focuses only on the expressions in the merged class, is trained to separate the images of the merged class into the prototypic expressions. The design of the 2-tiered structure allows us to use the appropriate features in each tier.

As illustrated in Fig.2, a hierarchical SVM classifier with two tiers is designed for the six prototypic expressions recognition. Firstly, we merged two of the six prototypic expressions (anger and sadness), which are the most commonly confused expressions, into one class. Together with the remaining four prototypic expressions, there are 5 classes in the first-tier classification. A 5-classes SVM classifier is used for this tier. The performance of the first-tier classification should be much better than directly classifying the six expressions since the major confusion has been removed by merging anger and sadness images together. For the images categorized as the merged class (anger and sadness), a 2-classes SVM is trained in the second tier to separate them into anger and sadness.

C. Feature selection in each tier

The first-tier classification plays such an important role in the whole recognition procedure that it should perform as accurately as possible. We trained the 5-classes SVM classifier based on the LBP feature, the displacement feature and the combined LBP and displacement feature separately. The results show that the combined feature yields the best performance in the classification of the 5 classes (*4 prototypic expressions and one merged class*). Thus, the combined feature is selected to represent facial expressions in the first tier.

After the first tier, the merged class flows into the second-tier classification, in which the images are separated into anger and sadness. The features highly related to these two expressions are adopted because the second-tier SVM only focuses on the classification of anger and sadness. We choose the

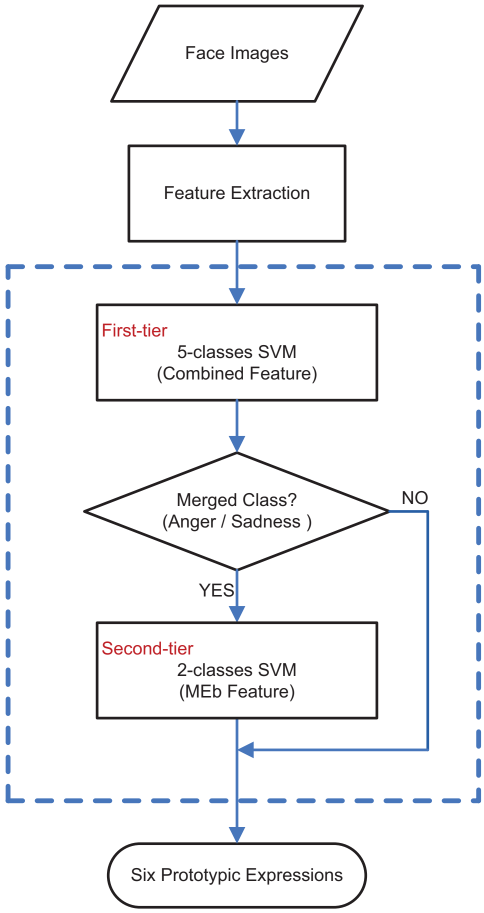

Fig. 2. The flow chart of the hierarchical classification.

displacement feature as the expressions representation at first, but approximately 20% sadness images are still misclassified as anger.

The inveterate confusion of anger and sadness evokes us to extract new feature since all the features used so far can not provide significant distinctions in representing these two expressions. Since the mouth and eyebrows possess the maximum amount of information related to the facial expressions[17], especially that the sadness is mainly conveyed by the mouth[18], we attempt to extract features of mouth and eyebrows to discriminate anger and sadness. As illustrated in Fig.1 (Right), the landmarks of mouth and eyebrows are selected and both the location and displacement of these landmarks are used to form the MEb (Mouth and Eyebrows) feature for the classification of anger and sadness in the second tier. The recognition result reveals that the selected MEb feature could separate anger and sad better than all the other features.

IV. Experiments

A. Experiment settings

Experiments have been conducted by applying the proposed method on the extended Cohn-Kanade (CK+) dataset [16],

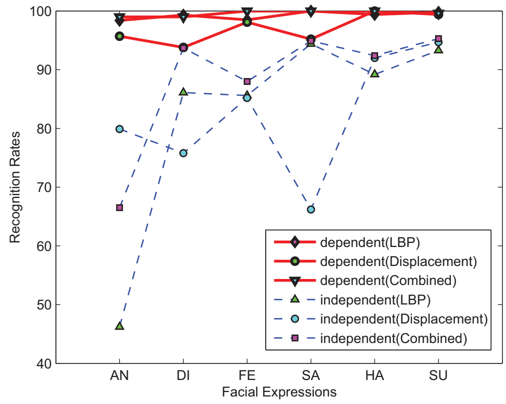

Fig. 3. The comparison of person-dependent and person-independent facial expression recognition.

which is comprised of 593 image sequences of 210 individuals. The expression in each sequence began with a neutral face and ended at the peak intensity. For all the 593 sequences, each image was AAM tracked with 68-points landmarks. However, only 327 of the 593 sequences carry the prototypic expression labels. The original images in the CK+ dataset are digitized into either 640×490 or 640×480 pixel arrays with 8-bit grey scale or 24-bit colour values.

In our experiments, only the images from the labelled 327 sequences are chosen to test the proposed method. The first 5 images of each sequence are ignored due to their low expression intensity. The selected images are aligned and resized into 110×150 pixel grey scale arrays according to the location of eyes and mouth and then split into training images and testing images.

B. Person-dependent vs person-independent

There are two ways to generate training set and testing set for a facial expression recognition system. One is person-dependent, while the other is person-independent. In the person-dependent case, the individuals included in the testing images also show up in the training images. It means that the classifier has seen the individuals included in the testing images. However, in the person-independent case, the individuals included in the testing images never appear in the training images. In this section, both the person-dependent and the person-independent cases are conducted on the selected images from CK+ dataset. The LBP features, displacement features and combined features are extracted as the expression representation.

Fig. 3 illustrates the performances of the person-dependent and person-independent expression recognition based on the LBP feature, displacement feature and the combined feature. The comparison shows that the confusions are very little in the person-dependent case, especially based on the combined feature. However, in the person-independent case, the expression recognition problem becomes much more difficult. The classification performances of the six prototypic expressions

	AN	DI	FE	SA	HA	SU
AN	**46.2**	2.5	0.4	**47.5**	1.5	1.9
DI	1.6	**86.1**	2.4	9.1	0.0	0.8
FE	0.0	3.7	**85.6**	10.2	0.0	0.5
SA	1.5	1.5	1.5	**94.4**	0.0	1.0
HA	0.8	1.3	6.1	1.3	**89.2**	1.3
SU	0.2	1.2	1.6	3.5	0.2	**93.3**

TABLE I
CONFUSION MATRIX OF PERSON-INDEPENDENT RECOGNITION BASED ON THE LBP FEATURE.

	AN	DI	FE	SA	HA	SU
AN	**79.9**	6.8	0.0	11.7	0.0	1.7
DI	9.2	**75.8**	0.0	15.1	0.0	0.0
FE	1.4	0.0	**85.2**	13.0	0.0	0.5
SA	**31.3**	0.0	2.0	**66.2**	0.0	0.5
HA	0.0	1.1	0.6	6.3	**92.0**	0.0
SU	0.0	0.9	0.9	2.6	0.9	**94.7**

TABLE II
CONFUSION MATRIX OF PERSON-INDEPENDENT RECOGNITION BASED ON THE DISPLACEMENT FEATURE.

decrease significantly, especially in the recognition of anger and sadness. The confusion matrices of person-independent recognition based on the LBP feature and displacement feature are recorded in Table I and Table II respectively. It can be seen that 47.5% of the anger images are misclassified as sadness in the LBP feature based recognition while 31.3% of the sadness images are confused as anger in the displacement feature based classification. Even based on the combined feature, the confusion of anger and sadness is still as high as 25.0% (as shown in Table III).

C. Results of the proposed method

The proposed hierarchical classification only focuses on the difficult person-independent expression recognition. In the proposed method, a 5-classes SVM classifier is trained in the first tier classification since anger and sadness have been merged into one class. The recognition performances based on the LBP feature, displacement feature and combined feature are illustrated in Fig. 4. Obviously, the combined feature yields the best classification performance. Thus, it is selected as the expression representation in the first tier.

The first-tier classification categorizes the images into 4 prototypic expressions plus the merged class. In order to finish the prototypic expression recognition, the images in the merged class need to be separated into anger and sadness in the second tier. We first utilize the displacement feature in the second-tier classification. The confusion matrix of the hierarchical classification is recorded in Table IV. Compared to the result in Table III, the hierarchical classification improves the recognition rate of anger from 66.5% to 86.0%. However, the recognition rate of sadness goes down to 75.8%, and the major confusion still lies between anger and sadness, with 21.7% of the sadness images misclassified as anger.

In order to better distinguish anger and sadness, we extract the MEb feature to be used in the second tier. Table V records the recognition result. It can be seen that the confusion between anger and sadness decreases to 14.6% while the sadness recognition rate reaches 93.4%. Although the rate for anger drops to 77.8%, the overall recognition rate reaches 89.2%, which shows that the selected MEb feature could separate anger and sad better than all the other features. Finally, we compare the proposed method with several state-of-art methods in table VI. It shows that the proposed method achieves the best performance in the recognition of anger, fear and sadness.

	AN	DI	FE	SA	HA	SU
AN	**66.5**	7.8	0.0	**25.0**	0.0	0.6
DI	2.0	**93.7**	0.0	4.0	0.0	0.4
FE	3.2	2.8	**88.0**	6.0	0.0	0.0
SA	0.5	2.5	0.5	**94.9**	0.0	1.5
HA	0.6	1.9	3.2	0.6	**92.4**	1.3
SU	0.0	0.9	1.6	2.1	0.0	**95.3**

TABLE III
CONFUSION MATRIX OF PERSON-INDEPENDENT RECOGNITION BASED ON THE COMBINED FEATURE.

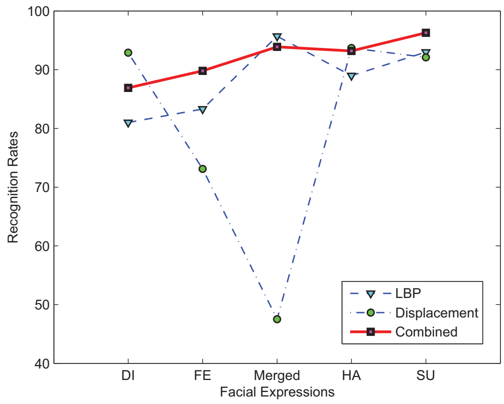

Fig. 4. The first-tier feature selection.

	AN	DI	FE	SA	HA	SU
AN	**86.0**	7.6	0.0	6.4	0.0	0.0
DI	6.7	**86.9**	0.0	6.0	0.0	0.4
FE	1.4	2.8	**89.8**	6.0	0.0	0.0
SA	21.7	0.0	0.5	**75.8**	0.0	2.0
HA	0.2	0.8	2.1	3.0	**93.2**	0.6
SU	0.5	0.0	0.7	2.6	0.0	**96.3**

TABLE IV
THE CONFUSION MATRIX OF THE HIERARCHICAL CLASSIFICATION (2-TIER BASED ON THE DISPLACEMENT FEATURE).

V. CONCLUSION

In this paper, a hierarchical classification approach is proposed for person-independent facial expression recognition. Due to the difficulty in distinguishing anger and sadness, they are combined into one class and join the other four prototypic

	AN	DI	FE	SA	HA	SU
AN	**77.8**	7.6	0.0	14.6	0.0	0.0
DI	1.6	**86.9**	0.0	11.1	0.0	0.4
FE	2.8	2.8	**89.8**	4.6	0.0	0.0
SA	4.0	0.0	0.5	**93.4**	0.0	2.0
HA	0.2	0.8	2.1	3.0	**93.2**	0.6
SU	0.7	0.0	0.7	2.3	0.0	**96.3**

TABLE V
THE CONFUSION MATRIX OF THE HIERARCHICAL CLASSIFICATION (2-TIER BASED ON THE MEB FEATURE).

Methods	AN	DI	FE	SA	HA	SU
P. Lucey [16]	75.0	94.7	65.2	68.0	**100.0**	96.0
O. Rudovic [23]	71.3	90.8	79.0	90.5	92.6	96.6
W. Gu [24]	75.3	**95.9**	85.5	90.0	93.3	97.7
L. Zhong [25]	71.4	95.3	81.1	88.0	95.4	**98.3**
The Proposed	**77.8**	86.9	**89.8**	**93.4**	93.2	96.3

TABLE VI
THE COMPARISON WITH THE STATE-OF-ART METHODS.

expressions in the first tier of classifications and then separated in the second tier. The hierarchical structure of the proposed method provides us with the opportunity to fuse different kinds of feature into the classification, which can enhance the recognition performance.

The experiment results on CK+ dataset show that the hierarchical SVM classifier improves the recognition performance for facial expression significantly, especially in reducing confusions between anger and sadness. We only test the proposed method on CK+ dataset because it is the only available dataset with facial landmark information. It is interesting that the selected mouth and eyebrow feature separates anger and sadness better than the displacement feature. This suggests that discriminative information of the prototypic expressions is conveyed by different facial components. Exploring this property further is a good way to refine facial expression systems.

REFERENCES

[1] M. Pantic and L. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 12, pp. 1424–1445, 2000.

[2] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[3] I. Cohen, N. Sebe, A. Garg, L. Chen, and T. Huang, "Facial expression recognition from video sequences: temporal and static modeling," *Computer Vision and Image Understanding*, vol. 91, no. 1, pp. 160–187, 2003.

[4] P. Ekman and W. Friesen, "Constants across cultures in the face and emotion." *Journal of personality and social psychology*, vol. 17, no. 2, p. 124, 1971.

[5] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *Image Processing, IEEE Transactions on*, vol. 16, no. 1, pp. 172–187, 2007.

[6] I. Kotsia, I. Buciu, and I. Pitas, "An analysis of facial expression recognition under partial facial image occlusion," *Image and Vision Computing*, vol. 26, no. 7, pp. 1052–1067, 2008.

[7] P. Michel and R. El Kaliouby, "Real time facial expression recognition in video using support vector machines," in *Proceedings of the 5th international conference on Multimodal interfaces*. ACM, 2003, pp. 258–264.

[8] P. Aleksic and A. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream hmms," *Information Forensics and Security, IEEE Transactions on*, vol. 1, no. 1, pp. 3–11, 2006.

[9] N. Sebe, M. Lew, Y. Sun, I. Cohen, T. Gevers, and T. Huang, "Authentic facial expression analysis," *Image and Vision Computing*, vol. 25, no. 12, pp. 1856–1863, 2007.

[10] H. Gu and Q. Ji, "Facial event classification with task oriented dynamic bayesian network," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2. IEEE, 2004, pp. II–870.

[11] M. Valstar, I. Patras, and M. Pantic, "Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*. IEEE, 2005, pp. 76–76.

[12] M. Bartlett, G. Littlewort, I. Fasel, and J. Movellan, "Real time face detection and facial expression recognition: Development and applications to human computer interaction." in *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*, vol. 5. IEEE, 2003, pp. 53–53.

[13] C. Shan, S. Gong, and P. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[14] Y. Tian, T. Kanade, and J. Cohn, "Evaluation of gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. IEEE, 2002, pp. 229–234.

[15] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*. IEEE, 1998, pp. 454–459.

[16] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[17] M. Pardàs and A. Bonafonte, "Facial animation parameters extraction and expression recognition using hidden markov models," *Signal Processing: Image Communication*, vol. 17, no. 9, pp. 675–688, 2002.

[18] F. Bourel, C. Chibelushi, and A. Low, "Recognition of facial expressions in the presence of occlusion," in *Proc. of the Twelfth British Machine Vision Conference*, vol. 1. Citeseer, 2001, pp. 213–222.

[19] Y. Tian, T. Kanade, and J. Cohn, "Recognizing action units for facial expression analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 2, pp. 97–115, 2001.

[20] M. Pantic and L. Rothkrantz, "Facial action recognition for facial expression analysis from static face images," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 34, no. 3, pp. 1449–1461, 2004.

[21] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

[22] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[23] O. Rudovic, V. Pavlovic, and M. Pantic, "Multi-output laplacian dynamic ordinal regression for facial expression recognition and intensity estimation," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2634–2641.

[24] W. Gu, C. Xiang, Y. Venkatesh, D. Huang, and H. Lin, "Facial expression recognition using radial encoding of local gabor features and classifier synthesis," *Pattern Recognition*, vol. 45, no. 1, pp. 80–91, 2012.

[25] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. Metaxas, "Learning active facial patches for expression analysis," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2562–2569.