

Boosting Coded Dynamic Features for Facial Action Units and Facial Expression Recognition

Peng Yang¹

Qingshan Liu^{1,2}

Dimitris N. Metaxas¹

¹Computer Science Department, Rutgers University
110 Frelinghuysen Road, Piscataway, NJ 08854, USA

²National Laboratory of Pattern Recognition, Chinese Academy of Sciences
Beijing, 100080, China

peyang@cs.rutgers.edu, qslu@cs.rutgers.edu, dnm@cs.rutgers.edu

Abstract

It is well known that how to extract dynamical features is a key issue for video based face analysis. In this paper, we present a novel approach of facial action units (AU) and expression recognition based on coded dynamical features. In order to capture the dynamical characteristics of facial events, we design the dynamical haar-like features to represent the temporal variations of facial events. Inspired by the binary pattern coding, we further encode the dynamic haar-like features into binary pattern features, which are useful to construct weak classifiers for boosting learning. Finally the Adaboost is performed to learn a set of discriminating coded dynamic features for facial active units and expression recognition. Experiments on the CMU expression database and our own facial AU database show its encouraging performance.

1. Introduction

Automatic facial expression and facial AU recognition have attracted much attention in the recent years due to its potential applications. Some methods [12] [1] [10] using static features in facial expression recognition have gotten some success, and all these methods assumed that facial expressions are static and recognition is done frame by frame in image sequence without using the temporal information. Actually, a natural facial event evolves over time from the onset, the apex and the offset, including facial expression and AUs. Therefore, static image based expression and AU recognition could not achieve a good performance in a practical system. The video-based methods using dynamic features get better performance [20], so dynamic features should be the optimal features to describe

facial expression and facial AUs [2] [18] [3]. Although much progress has been made by lots of methods [3] [13], achieving high accuracy is still a challenge due to the complexity and variation of facial dynamics. Normally there are two common approaches to extract facial features: geometric feature-based methods [6] [7] [15] and appearance-based methods [1] [12]. Appearance features have been demonstrated to be better than geometric features [19] [14], because geometric feature-based methods are sensitive to inaccurate image alignment and motion discontinuities. Gabor appearance feature has been widely adopted to describe local appearance [8] [1] [4], but its computation expense is much higher than other local features such as local binary pattern (LBP) [9] and haar features [16].

In this paper, we propose a new approach of facial active units (AU) and expression recognition. In order to capture the temporal characteristic of facial AU and expression, we design dynamic haar-like features to represent the facial images. Compared to Gabor features, haar-like features are very simpler than Gabor representation in computation cost, since they are just based on simple add or minus operators [16]. The dynamic haar-like features are coded into binary patterns for further effective analysis inspired by [9]. Finally based on the coded dynamic features, the Adaboost is employed to learn the combination of optimal discriminate features to construct the classifier. The experiments carried on two databases show the promising performance of our proposed method.

The paper is organized as following, the overview of our proposed method is shown in section 2; in section 3, we introduce the dynamic features which we use to describe the facial dynamics. We describe how to create the code book and encode the dynamic features in section 4, followed by the experiment part with the comparison with other approach using the static features.

2. Overview of the Proposed Approach

Since facial AU and expression variations are dynamical in the temporal domain, using the variations of temporal information for facial AU and expression recognition is a trend [5]. In this paper, we propose a novel framework for facial AU and expression recognition based on coded dynamical features. Figure 1 illustrates the structure of our framework, which has three main components: dynamical feature extraction, coding dynamical features, and Adaboosting learning. For the component of dynamical feature extraction, we design dynamical Harr-like features to capture the temporal variations of facial AUs and expressions. Inspired by the binary pattern coding [9], we analyze the distribution of each dynamical haar-like feature, and we create a code book for it. According to the code books, the dynamical haar-like features are further mapped into binary pattern features. Finally the Adaboost is used to learn a set of discriminating coded features for facial AU and expression recognition.

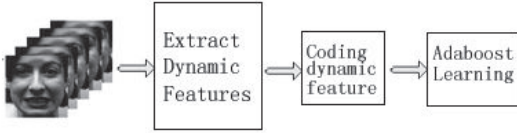


Figure 1. The structure of the proposed framework.

3. Dynamical Haar-like Feature Representation

Haar-like feature achieved a great performance in face detection [16]. Considering its much lower computation expense compared with Gabor features, we exploit the haar-like features to represent face images, and extend it to represent the dynamic characteristic of facial expression and AU. Figure 3 shows the example of extracting haar-like features in a face image. The dynamical Harr-like features are built by two steps: (1) Thousands of haar-like features are extracted in each image frame. (2) The same haar-like features in the consecutive image frames are combined as the dynamic features. Figure 2 shows the flowchart of dynamical Harr-like features, and the details are described in the following.

For simplicity, we denote one image sequence I with n frames, and each frame with the label I_i , where i is the index of the frame. We set H as the haar-like feature set which includes all the haar-like features in one face image. For each haar feature in frame I_i , we give it label $h_{i,j}$, where i is the index of the frame and j is the index of the haar-like feature in the feature set H .

Based on each haar-like feature $h_{i,j}$, we build a dynamical Harr-like feature $u_{i,j}$ in the temporal domain as $u_{i,j} = \{h_{i-k,j}, h_{i-k+1,j}, \dots, h_{i+k,j}\}$. Figure 4 shows an illustration. We call each $u_{i,j}$ as a dynamical feature unit. The temporal variation of facial AUs and expressions can be effectively described by all the all the $u_{i,j}$ s.



Figure 3. Example of Haar-like features superimposed onto a face image

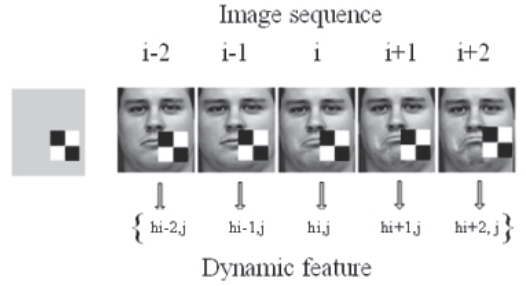


Figure 4. Example of one dynamic feature unit $u_{i,j}$ in an image sequence

4. Coding the Dynamic Features

As mentioned above, a dynamic feature unit is composed of a set of haar-like features in the same position. Thus, each dynamical unit is a feature vector. In this paper, we further code a dynamic feature unit into a binary pattern inspired by [4] [9]. This coding scheme has two advantages: 1) Constructing weak learner for Ababoost learning with one feature is easier than with a feature vector; 2) the proposed binary coding is based on statistical distribution of training samples, so it is robust to noise.

We create the code books for each facial expression and each AU based on the training samples. First, we analyze the distribution of each feature unit u_j under each expression or each AU, and then the mean μ and the variance σ can be estimated from this distribution. The gaussian distribution $N_j(\mu, \sigma)$ is adopted to estimate the distribution of the feature unit u_j . Then we obtain the code book $B_k(N_1(\mu_1, \sigma_1), N_2(\mu_2, \sigma_2), \dots, N_m(\mu_m, \sigma_m))$ for all of them under corresponding expression or AU, where k is the label of the different expressions or AUs, j is the index of the feature unit and m is the size of the feature set H . Based on the code book, we can map the each dynamical haar-like feature $h_{i,j}$ to $\{1, 0\}$ pattern by the formula 1:

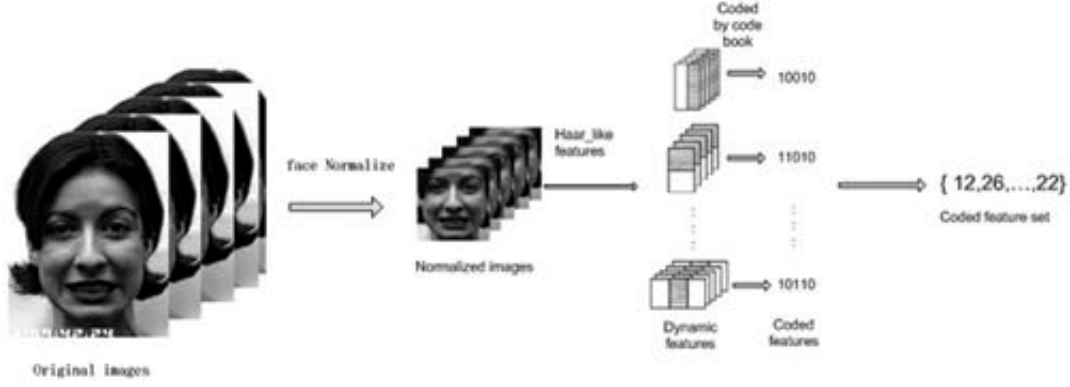


Figure 2. The flowchart of dynamic features extraction

$$C_{i,j} = \begin{cases} 0 & : \text{if } \frac{\|h_{i,j} - \mu_j\|}{\sigma_j} > T \\ 1 & : \text{if } \frac{\|h_{i,j} - \mu_j\|}{\sigma_j} < T \end{cases} \quad (1)$$

where T is the threshold, because the standard normal gaussian distribution $x \sim N(0,1)$, $Pr(\|x\| \leq 1.65) = 95\%$, therefore we set $T = 1.65$ in our experiments to cover almost all the positive example.

Based on the formula 1, we can map one haar-like dynamic feature unit $u_{i,j}$ to one binary pattern $T_{i,j}$. Figure 5 gives the procedure of creating the coded feature T .

$$T_{i,j} = \{C_{i-L,j}, C_{i-L+1,j}, \dots, C_{i,j}, \dots, C_{i+L-1,j}, C_{i+L,j}\} \quad (2)$$

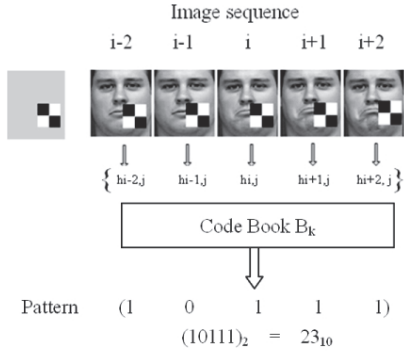


Figure 5. Example of coding one dynamical feature unit $u_{i,j}$

5. Boosting the Coded Dynamic Features

Since there exist thousands of dynamical coded features, a set of discriminating features should be selected to construct the final classifier. In this paper, we use the Adaboost learning to achieve this goal, and the weak learner is built on each coded feature as in [16]. For each expression, we set the image sequences of this expression as the positive

examples, and the image sequences of other expressions as the negative samples. Therefore, one classifier for each expression can be built based on the corresponding coded features. The learning procedure is in Figure 6. The testing procedure is displayed in the Figure 7. Our expression performed a 6-way forced choice between the following emotion categories: Happiness, sadness, surprise, disgust, fear, anger, AdaBoost performed binary decision tasks. Six Adaboosts were trained to discriminate each emotion from other things.

1. Given example image sequences $(x_i, y_i), \dots, (x_n, y_n)$, $y_i \in \{1, 0\}$ for specified expression and other expressions respectively.
2. Initialize weight $D_t(i) = 1/N$.
3. Get the dynamic features on each image sequence.
4. Code the dynamic features based on the corresponding code book, and get $T_{i,j}$. Build one weak classifier on each coded feature.
5. Use standard Adaboost to get strong classifier $H(x_i)$.

Figure 6. Learning procedure based on AdaBoost.

6. Experiments

To evaluate the performance of the coded dynamic features, two databases are used in our experiment. The first one is the CMU Cohn-Kanade facial expression database, and another one is the our facial AU database. Psychophysical studies indicates that basic emotions have corresponding universal facial expressions across all cultures, and there is a set of prototypic emotional expressions such as disgust, fear, happiness, surprise, sadness and anger. These six expressions have the strong correlation with just a set of AUs, therefore we just do some experiments on those AUs for AU recognition. In our experiments, we use Viola's [16] method to detect face and normalize the face automatically based on the location of the eyes.

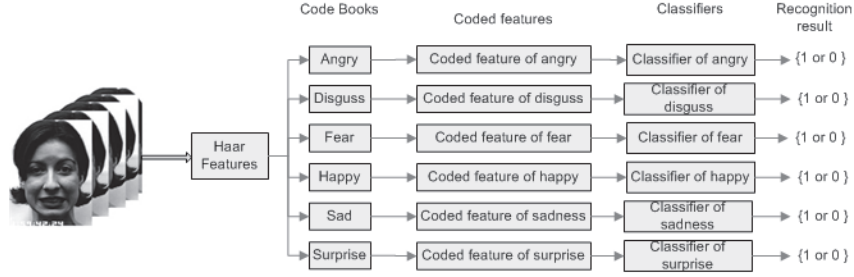


Figure 7. The procedure of the expression recognition

6.1. Experiment on Facial Expressions

The Cohn-Kanade Facial Expression Database[9] consists of 100 university students aged from 18 to 30 years, of which 65% are female, 15% are African-American, and 3% are Asian or Latino. Subjects are instructed to perform a series of 23 facial displays, six of which are prototypic emotions mentioned above. For our experiments, we select 300 image sequences from the database. The selection criterion is that a sequence could be labeled as one of the six basic emotions. The sequences come from 96 subjects, with 1 to 6 emotions per subject. For each sequence, the 8~12 frames previous peak are used, we randomly select 60 subjects as the training set, and the rest subjects as the testing set. Face image is normalized to 64×64 based on the experiment result of Tian [13], the examples are shown in Figure. 8.



Figure 8. Examples of six basic expressions.(Anger, Disgust, Fear, Happiness, Sadness and Surprise

On the training set, we analyze the distribution of feature units under each expression, therefore we get 6 code books for 6 expressions. According to the specified code book, we get the binary sequence for each subject under the corresponding expression. In this experiment, we set the coded feature as

$$T_{i,j} = \{C_{i-2,j}, C_{i,j}, C_{i+2,j}\} \quad (3)$$

and ignore the previous frame $i - 1$ and posteriors frame $i + 1$ because there is almost no difference between two neighbored frames.

In [12] [17], they gave some recognition ratio on each expression or AU, however, we can not see the false alarm from their results. To evaluate the real performance of our method, ROC curve is used in our experiment to evaluate the performance. In order to show the advantage of the coded dynamic feature, we use the Adaboost and static Haar

Table 1. The Area under the ROC curves (Expression)

Expression	Static feature + AdaBoost	Coded dynamic feature + AdaBoost
Angry	0.856	0.973
Disgust	0.898	0.941
Fear	0.841	0.916
Happiness	0.951	0.9913
Sadness	0.917	0.978
Surprise	0.974	0.998

features in each image frame on the same training and testing data set. Figure. 9 reports the results, and we can see the ROC curves on the coded dynamic feature are much better the ones on the static feature. And the area under the ROC curves is listed in the table 1.

6.2. Experiment on Facial AUs

According to Ekman's Group description method of Facial Action Coding System (FACS) [11], totally there are 71 primitive units, called Action Unit(AU). Based on them, any expression display can be represented by a single AU or a combination of several AUs. We test our approach on the our facial AUs database, this data set has spontaneous facial expressions from freely behaving individuals and it is labeled by psychological experts. The data set consisted of 300 Gigabytes of 640×480 color images, 8 bits per pixels, 60 fields per second, 2:1 interlaced. We have 33 subjects in current data set. Because there are few samples for some AUs, in our experiments, we just focus on 8 AUs: AU1, AU2, AU4, AU5, AU10, AU12, AU14 and AU20. The examples are shown in Figure. 10. We randomly select 22 subjects as the training set, and the other 11 subject as the testing set. Similar to the experiment on the facial expression recognition, we analyze the distributions of feature units under each AU, and get 8 code books for 8 AUs. We use the same code feature $T_{i,j}$ as in (3). To show the advantage and the power of the coded dynamic features, we use the static haar features in a single frame, training is done

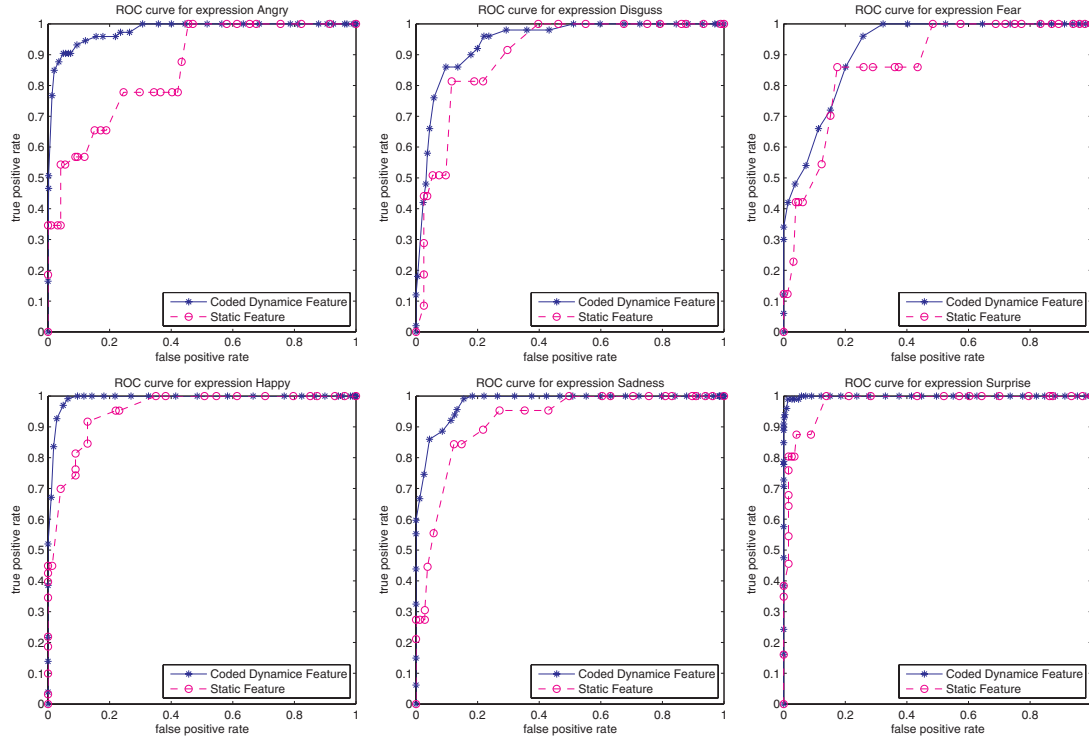


Figure 9. ROC curves of expression based on coded dynamic features and static features

on the same training set and test on the same testing set. The ROC curves are reported in Figure. 11, we can clearly see the coded dynamic feature is much better than the static feature. And the area under the ROC curves is listed in the table 2.



Figure 10. ROC curves of expressions based on coded dynamic features and static features

7. Conclusions

This paper presented a novel approach for video based facial AU and expression recognition, which is based on coded dynamical features. We first extracted dynamical Haar-like features to capture the temporal information of facial AUs and expressions, and then further coded them into binary pattern features inspired by the binary pattern coding. Finally the Adaboost was performed to learn a set

Table 2. The Area under the ROC curves (AUs)

AUs	Static feature + AdaBoost	Coded dynamic feature + AdaBoost
AU1	0.718	0.7715
AU2	0.714	0.784
AU4	0.606	0.8213
AU5	0.652	0.7386
AU10	0.660	0.663
AU12	0.850	0.8624
AU14	0.735	0.7748
AU20	0.601	0.7163

of discriminating coded features to construct the final classifier. Experiments on the CMU facial expression database and our own facial AU database showed that the proposed method has a promising performance. In addition, this method can be easily extended to video based face recognition.

References

- [1] M. Bartlett, G. Littlewort, I. Fasel, and J. Movellan. Real time face detection and facial expression recognition: Development and applications to human computer interaction. *Proc. of Intl Conf. CVPR Workshop on Computer Vision and*

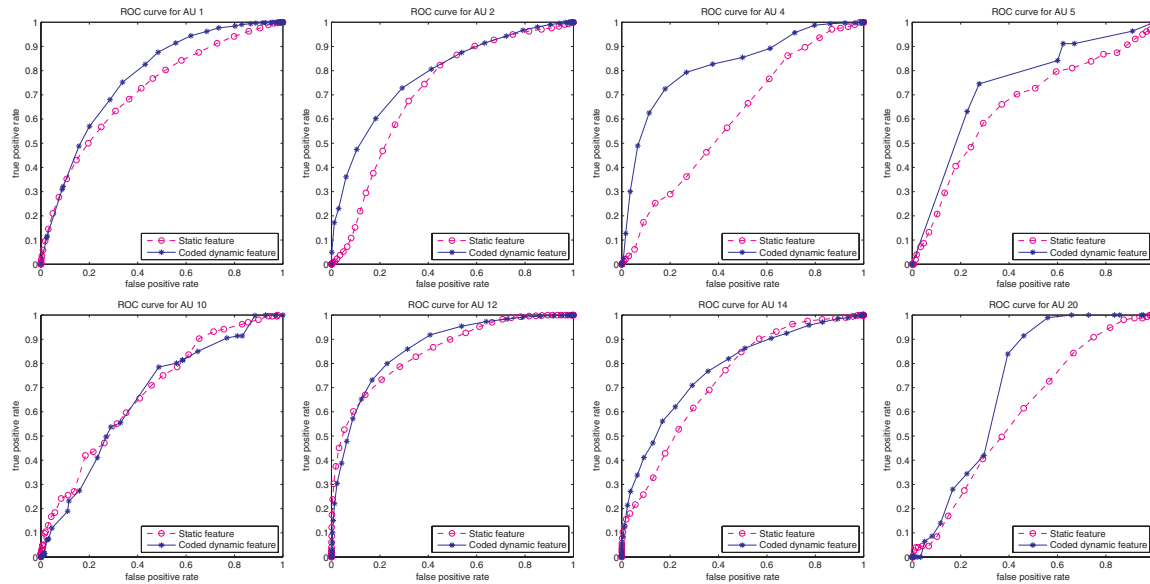


Figure 11. ROC curves of AUs based on coded dynamic features and static features

Pattern Recognition for Human-Computer Interaction, 2003.

1

- [2] M. J. Black and Y. Yacoob. Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Journal of Computer Vision*, 1997. 1
- [3] I. Cohen, N. Sebe, L. Chen, A. Garg, and T. Huang. Facial expression recognition from video sequences Temporal and static modeling. *Computer Vision and Image Understanding*, 2003. 1
- [4] J. Daugman. Demodulation by complex-valued wavelets for stochastic pattern recognition. *Int'l J. Wavelets, Multiresolution and Information Processing*, 2003. 1, 2
- [5] Z. G. and P. M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007. 2
- [6] H. Gu and Q. Ji. Facial event classification with task oriented dynamic bayesian network. *Proc. of Intl Conf. Computer Vision and Pattern Recognition*, 2004. 1
- [7] J.-J. J. Lien, T. Kanade, J. Cohn, and C. Li. Detection, tracking, and classification of action units in facial expression. *Journal of Robotics and Autonomous Systems*, 1999. 1
- [8] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 2002. 1
- [9] T. Ojala and M. Pietikainen. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002. 1, 2
- [10] M. Pantic and J. Rothkrantz. Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man and Cybernetics*, 2004. 1
- [11] P. Ekman, W. V. Friesen, and J. C. Hager. Facial action coding system: manual. 2002. 4
- [12] C. Shan, S. Gong, and P. W. McOwan. Conditional mutual information based boosting for facial expression recognition. *British Machine Vision Conference*, 2005. 1, 4
- [13] Y. Tian. Evaluation of face resolution for expression analysis. *Proc. of Intl Conf. CVPR Workshop on Face Processing in Video (FPIV'04)*, 2004. 1, 4
- [14] Y. Tian, T. Kanade, and J. Cohn. Evaluation of gabor wavelet-based facial action unit recognition in image sequences of increasing complexity. *Proc. of Intl Conf. Automatic Face and Gesture Recognition (FG'02)*, 2002. 1
- [15] M. F. Valstar, I. Patras, and M. Pantic. Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data. *Proc. of Intl Conf. CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction*, 2005. 1
- [16] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 2002. 1, 2, 3
- [17] J. Whitehill and C. W. Omlin. Haar features for faces au recognition. *Proc. of Intl Conf. Automatic Face and Gesture Recognition*, 2006. 4
- [18] Y. Yacoob and L. Davis. Computing spatio-temporal representations of human faces. *Proc. of Intl Conf. Computer Vision and Pattern Recognition*, 1994. 1
- [19] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu. Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. *Proc. of Intl Conf. Automatic Face and Gesture Recognition*, 1998. 1
- [20] G. Zhao and M. Pietik0Elinen. Dynamic texture recognition using volume local binary patterns. *Proc. of European Conference on Computer Vision*, 2006. 1