Recognizing Facial Expressions at Low Resolution

Caifeng Shan, Shaogang Gong and Peter W. McOwan

Deparment of Computer Science, Queen Mary, University of London, London, E1 4NS, UK

{cfshan, sgg, pmco}@dcs.qmul.ac.uk

Abstract

This paper focuses on recognizing facial expressions at low resolution. We introduce Local Binary Patterns (LBP) as novel low-computation discriminative features for low-resolution facial expression recognition. Compared to Gabor wavelets, LBP features can be derived rapidly in a single scan of raw images, whilst still retaining enough facial infomation in a compact representation. Support Vector Machine (SVM) is adopted to classify facial expressions. Extensive experiments on the Cohn-Kanade database demonstrate that the LBP features are effective and efficient for facial expression recognition, and crucially perform robustly and stably over a useful range of low resolutions. Our method yields promising performance when processing compressed low-resolution video sequences from the PETS 2003 dataset.

1. Introduction

Facial expression is one of the most powerful, nature, and immediate means for humans to communicate their emotions and intentions [1]. Due to its potential applications, such as human-computer interaction and visual surveillance, automatic facial expression recognition has attracted much attention [2, 3] over the past two decades, and much progress has been made [4, 5, 6, 7, 8, 9, 10, 11].

Automatic facial expression recognition involves both facial feature representation and classifier design. Facial feature representation is to derive a set of features from original face images which minimizes the intro-class variations of expressions and maxmizes the extra-class ones. If inadequate features are used, even the best classifier could fail to realize accurate recognition. There are two main types of approaches to extract facial features: geometric feature-based methods and appearance-based methods [12]. Geometric features present the shape and locations of facial components, which are extracted to form a feature vector that represents the face geometry. With appearance-based methods, image filters, such as Gabor wavelets, are applied either to the whole-face or to specific face-regions. Zhang [13] compared two types of facial features: the geometric positions of 34 fiducial points on a face and a set of multi-

scale, multi-orientation Gabor wavelet coefficients at these points for expression recognition; Gabor-wavelet representation was much effective. Donato et al [7] explored different techniques to face image representation for facial action recognition, which include holistic spatial analysis, such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Local Feature Analysis (LFA), and Linear Discriminant Analysis (LDA); and local schemes such as Gabor-wavelet representation and local principal components; Gabor-wavelet representation and ICA performed best. Recently Buciu et al [14] adopted ICA and Gabor-wavelet representation for facial expression recognition. Although Gabor-wavelet representations have been widely adopted [13, 7, 15, 14, 16], it is computationally expensive to convolve face images with multi-banks of Gabor filters to extract multiscale and orientational coefficients.

On classifier design, different classifiers have been applied to facial expression recognition, such as Neural Networks (NN) [13, 16], Support Vector Machines (SVM) [14, 17], Bayesian Networks (BN)[10], Linear Discriminant Analysis [6], and rule-based classifiers [8] for image-based expression recognition, or Hidden Markov Model (HMM) [10] and Dynamic Bayesian Networks (DBNs) [11] for sequence-based expression recognition. Cohen et al [10] compared Naive-Bayes classifiers, and Gaussian Tree-Augmented Naive (TAN) Bayes classifiers for expression recognition from static images. They further proposed a multi-level HMM classifier, which not only performs expression classification on a video segment, but also segments automatically an arbitrary long video sequence into different expressions segments without resorting to heuristics. Recently Bartlett et al [15, 17] conducted a systematic comparison of machine learning methods such as AdaBoost, SVM, and LDA applied to expression recognition.

One limitation of the existing methods is that they attempt to recognize facial expressions from data collected in a highly controlled environment given high resolution frontal faces [16]. However, in real-world environments such as smart meeting and visual surveillance, the input face images are often in lower resolution. Fig 1 gives a real-world image recorded in a smart meeting scenario. Obviously low-resolution images and low-intensity expressions in real-world environments make real-life expression recog-
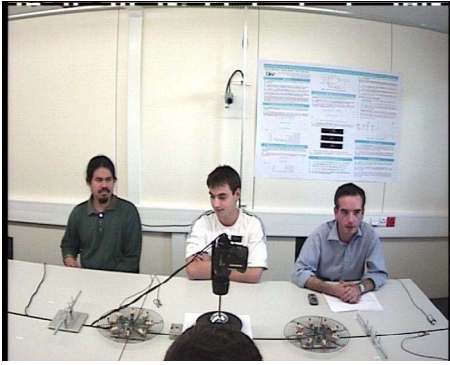
Figure 1: An example of low-resolution facial expressions recorded in real-world environments. From PETS 2003 dataset.

nition much more difficult. Recently Tian et al [1, 16] made a first attempt to recognize facial expressions with low resolution. They proposed a real-time system to recognize facial expression in compressed images with low resolution [1]. In [16], Tian further investigated the effects of different image resolutions for each step of automatic facial expression analysis.

In this paper, we focus on recognizing facial expression at low resolution. We introduce Local Binary Patterns (LBP) as novel low-cost discriminative appearance feature for low-resolution facial expression recognition. Local Binary Patterns were proposed originally for texture analysis [18, 19]. Recently Ahonen et al [20, 21] presented LBP based methods for face detection and recognition. Our motivation is that face images can be seen as a composition of micro-patterns which can be effectively described by LBP. Compared to Gabor wavelets, LBP features can be derived very fast in a single scan of raw images, whilst still retaining enough facial infomation in a compact representation. We adopt the Support Vector Machine (SVM) classifier as suggested elsewhere [14, 15, 17]. Extensive experiments on the Cohn-Kanade Facial Expression Database [22] demonstrate that the LBP features are effective and efficient for facial expression recognition. Experiments on face images of different resolutions show that the LBP features perform robustly and stably over a useful range of low resolutions. More crucially, our algorithm performs well on compressed low-resolution video sequences from the PETS 2003 dataset, which illustrate its suitability for real-world environments.

The paper is structured as follows. In Section 2, we introduce Local Binary Patterns. Section 3 discusses the classifier for facial expression recogniton. Section 4 presents extensive experiments on the Cohn-Kanade database and the PETS test sequences. We summarize in Section 5.

2. Local Binary Patterns

The original LBP operator was introduced by Ojala et al [18]. The operator labels the pixels of an image by thresholding a 3×3 neighbourhood of each pixel with the center value and considering the results as a binary number (See the left side of Fig 2). The 256-bin histogram of the labels computed over a region can be used as a texture descriptor. Each bin (LBP code) can be regarded as a micro-texton.
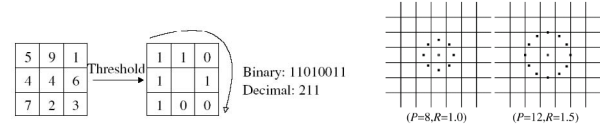

Figure 2: Left: The basic LBP operator [20]. Right: Two examples of the extended LBP [19]: a circular $(8, 1)$ neighborhood, and a circular $(12, 1.5)$ neighbourhood.

The limitation of the basic LBP opterator is its small 3×3 neighbourhood which can not capture dominant features with large scale structures. The operator was extended to use neighbourhood of different sizes [19]. Using circular neighbourhoods and bilinearly interpolating the pixel values allow any radius and number of pixles in the neighbourhood. Fig 2 gives two examples of the extended LBP operator, where the notation (P, R) denotes a neighborhood of P equally spaced sampling points on a circle of radius of R that form a circurlarly symmetric neighbour set.

The LBP operator $LBP_{P,R}$ produces 2^P different output values, corresponding to the 2^P different binary patterns that can be formed by the P pixels in the neighbour set. It has been shown that certain bins contain more information than others [19]. Ojala et al called these fundamental patterns uniform patterns. A Local Binary Pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 and vice versa when the binary string is considered circular. For example, 00000000, 001110000 and 11100001 are uniform patterns. It is observed that uniform patterns account for nearly 90% of all patterns in a $(8, 1)$ neighbourhood and for about 70% in a $(16, 2)$ neighbourhood in texture images [19]. Accumulating the patterns which have more than 2 trainsitions into a single bin yields an LBP operator, denoted $LBP_{P,R}^{u2}$, with less than 2^P bins. For example, the number of labels for a neighbrhood of 8 pixels is 256 for the standard LBP and 59 for the LBP^{u2}.

After labeling a image with the LBP operator, a histogram of the labelled image $f_l(x, y)$ can be defined as

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

where n is the number of different labels produced by the

LBP operator and

$$I(A) = \begin{cases} 1 & A \text{ is true} \\ 0 & A \text{ is false} \end{cases} \quad (2)$$

This LBP histogram contains information about the distribution of the local micro-patterns, such as edges, spots and flat areas, over the whole image.

Face images can be seen as a composition of micro-patterns which can be effectively described by the LBP features. However, a LBP histogram computed over the whole face image encodes only the occurrences of the micro-patterns without any indication about their locations. To also consider shape infomation of faces, face images are first divided into small regions $R_0, R_1, \ldots, R_m$ to extract LBP features (See Fig 3 for an illustration). The LBP features extracted from each sub-region are concatenated into a single, spatially enhanced feature histogram defined as:

$$H_{i,j} = \sum_{x,y} I\{f_l(x,y) = i\} I\{(x,y) \in R_j\} \quad (3)$$

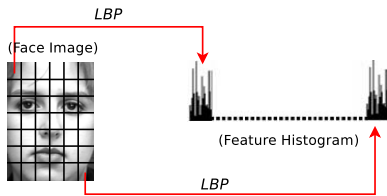where $i = 0, \ldots, n-1, j = 0, \ldots, m-1$.

Figure 3: A face image is divided into small regions from which LBP histograms are extracted and concatenated into a single, spatially enhanced feature histogram.

The extracted feature histogram represents effectively the local texture and global shape of face images. In this histogram, the face is described at three different levels of locality: the labels for the histogram contain the pixel-level patterns, the labels are summed over a small region to produce information on a regional level and the regional histograms are concatenated to build a global description of the face [20].

3. Facial Expression Recognition

We adopt the Support Vector Machine classifier [23] for facial expression recognition. SVM performs an implicit mapping of data into a higher dimensional feature space, where linear algebra and geometry can be used to separate data that is only separable with nonlinear rules in input space.

Given a training set of labeled examples $T = \{(x_i, y_i), i = 1, \ldots, l\}$ where $x_i \in R^n$ and $y_i \in \{1, -1\}$,

the new test data x is classified by the following function:

$$f(x) = sgn\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b\right) \quad (4)$$

where α_i are Lagrange multipliers of a dual optimization problem, $K(x_i, x)$ is a kernel function, and b is the parameter of the optimal hyperplane. It is possilbe to show that only some of the α_i are non-zero in the optimal solution. Given a non-linear mapping Φ that embed input data into feature space, kernels have the form of $K(x_i, x_j) = \langle \Phi(x_i) \cdot \Phi(x_j) \rangle$. SVM finds a linear separating hyperplane with the maximal margin to separate the training data in feature space, and the hyperplane is defined by the type of kernel function used.

SVM allows domain-specific selection of the kernel function. Though new kernels are being proposed, in our experiment, we used linear, polynomial, and RBF kernels. SVM makes binary decisions, so multi-class classification here is accomplished by a cascade of binary classifiers together with a voting scheme.

4. Experiments

4.1 Experiments: Cohn-Kanade Database

The proposed recognition algorithm was first trained and tested on the Cohn-Kanade database [22]. This database consists of 100 university students in age from 18 to 30 years, of which 65% were female, 15% were African-American, and 3% were Asian or Latino. Subjects were instructed by an experimenter to perform a series of 23 facial displays, six of which were based on descriptions of prototypic emotional expressions such as anger, disgust, fear, joy, sadness, and surprise. Image sequences from neutral to target display were digitized into 640 by 490 pixel arrays with 8-bit precision for grayscale values.

We selected 320 image sequences from the database. The only selection criterion was that a sequence can 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 neutral face and three peak frames were selected for 7-class basic emotional expression recognition. To evaluate generalization performance, a 10-fold cross-validation test scheme was adopted.

Following to Tian [16], we normalized the faces to a fixed distance of 55 pixels between the centers of the two eyes. It is observed that the width of a face is roughly twice this distance, and the height is roughly triple. Hence, facial images of 110×150 pixels were cropped from orignial frames based on the two eyes location. No further alignment of facial features such as alignment of mouth [13] was performed in our algorithm. Due to LBP's gray-scale in-

variance, there was no attempt made to remove illumination changes (as Tian did in [16]) in our algorithm.

Some parameters can be optimized for the LBP feature selection. The first one is the LBP operator, and another is the number of regions divided. We selected the 59-bin $LBP_{8,2}^{u2}$ operator, and divided the 110×150 pixels facial images into 18×21 pixels regions, giving a good trade-off between recognition performance and feature vector length [20]. Thus facial images were divided into $42(6 \times 7)$ regions as shown in Fig 3, and represented by the LBP histograms with length of $2,478(59 \times 42)$.

We adopted the SVM implementation in SPIDER [1] to perform facial expression recognition. The generalization performance with linear, polynomial and RBP kernels are 87.2%, 88.4% and 87.6% respectively. Experimental results illustrate the proposed LBP-based method is effective for facial expression recognition.

	LBP + SVM	Gabor-wavelet + SVM [15]
Recogniton Rates (Linear)	87.2%	84.8%
Recogniton Rates (Polynomial)	88.4%	worse than RBF/Linear
Recogniton Rates (RBF)	87.6%	86.9%
Feature Dimension	2,478	92,160

Table 1: Comparisons between the Gabor-wavelets based SVM [15] and our LBP-based SVM.

Recently Bartlett et al [15] conducted similar experiments based on Gabor-wavelet representation. They selected 313 image sequences from the Cohn-Kanade database. The sequences came from 90 subjects, with 1 to 6 emotions per subject. The facial images were converted into a Gabor magnitude representation using a bank of 40 Gabor filters. They performed 10-fold cross-validation experiments using SVM with linear, polynomial, and RBF kernels. Linear and RBF kernels perfromed best, with recognition rates of 84.8% and 86.9% respectively; results of polynomial kernel were not given in their paper.

Comparisons summaried in Table 1 show that the performance of our LBP-based SVM method is superior to the Gabor-wavelet based SVM algorithm. More crucially though, our advangtage lies at the simplicity of LBP features allows very fast feature extraction, without complex analysis in extracting a large set of Gabor wavelet coefficients. Compared to the high dimensionality of the Gabor respsrentation $O(10^5)$ in [15], the LBP features with short vector length $O(10^3)$ lies in a significantly lower dimensional space.

4.2 Experiments: Effects of Resolution

Since our aim is low-resolution facial expression recognition, we further evaluated our method for low-resolution input images. Specially we investigated its performance on different face resolutions. As shown in Table 2, totally six different resolutions of the face region were studied (110×150, 55×75, 36×48, 27×37, 18×24, and 14×19 pixels) based on the Cohn-Kanade database. The lower resolution images were down-sampled from the original images.

For LBP feature extraction, lower resolution face images were divided into 10×10 pixels regions (which may overlap with each other in the small face images). We adopted the 4-neighborhood LBP operator $LBP_{4,1}$ for each sub-region, which has 16 lables.

To compare with Tian's work [16], we conducted a 10-fold cross-validation experiment on 6-class basic expression recognition. The general performance of our method on different resolutions are shown in Table 2.

	110×150	55×75	36×48	27×37	18×24	14×19
LBP	92.1%	88.6%	86.3%	83.5%	78.7%	75.8%
AP [16]	92.2%	91.6%	-	77.6%	-	68.2%
G1 [16]	91.8%	91.6%	-	N/A	-	N/A
G2 [16]	73.8%	72.9%	-	61.3%	-	N/A

Table 2: Comparisons on image resolutions between our LBP-based algorithm and Tian's methods [16].

Recently Tian investigated the effects of different image resolutions for each step of automatic facial expression analysis [16]. In her experiment, 375 image sequences were selected from the Cohn-Kanade database for 6-class emotion classification. Tian extracted two types of features: geometric features and apperance features. Geometric features were extracted by feature tracking (G1) [9] and feature detection (G2) [1] respectively. For appearance features (AP), a bank of 40 Gabor filters were applied to the difference image to extract facial appearance changes. The difference images were obtained by substracting a neural expression. Tian used a three-layer Neural Network to recognize expressions. Recognition results of Tian's methods are summarized in Table 2 [2]. An aspect should be noticed that Tian did not conduct cross-validation, but randomly separated the dataset into a training set and a test set.

We can draw the following conclusions from the experimental results: (1) Geometric features are not available for

---

lower resolution, while appearance features such as Gabor and LBP could be extracted on different resolutions. Since facial components such as mouth, eyes, brows and nose can not be correctly detected or tracked in lower resolution images, hence geometric features are not reliable in low-resolution images. On the contrary, appearance features present the apperance changes of the face such as wrinkles and furrows, so they are available even in lower resolution images. (2) The proposed LBP-based SVM algorithm is more effective for expression recognition with low resolution images than that the Gabor-wavelets based NN method. Additionally, neutral expressions were required when Tian extracted Gabor representation, while LBP feature extraction did not need neutral faces. (3) The LBP features perform robustly and stably over a useful range of low resolutions, which is critical in real-world applications where low-resolution video input is only available.

4.3   Expereriments: PETS Dataset

After verifying its effectiveness for low-resolution face images, we tested our method on compressed low-resolution image sequences recorded in a real environment. We used the smart meeting dataset in the PETS 2003 evaluation datasets. Results on scenario A, camera 1 were reported here. In this scenario, each person enters the conference room one after the other, goes to his place, presents himself to the frontal camera, and sit down. Then each person looks at the other people with different expressions. Fig 1 shows a frame in the video sequence. Three facial expressions, neutral, anger and joy, are available in the dataset.

The real-world video sequence contains the full range of head motion. In Tian's system [1], the head pose was first estimated based on the detected head; then for frontal and near frontal views of the face, the facial features were extracted for facial expression recognition. Since our focus was investigating the validity of our method in real-world compressed video inputs, we did not consider pose estimation currently. We cropped the face region in frontal and near frontal views based on the location of two eyes from the input image sequence, then performed facial expression recognition on the cropped face images. Fig. 4 shows face regions cropped in Frame 17130.

It is very difficult, even for human beings, to recognize the expressions at very low intensity. Similiar to Tian et al [1], experiments were conducted on showing some frames of expression in low intensity to a small set of human observers (in this instance five researchers in our lab) resulting in many who could not perform recognition against the ground truth provided by the PETS dataset (original GT). Tian et al modified the ground truth based on the majority. Here we also generated a new ground truth (modified GT) based on human observations. Some examples of modified
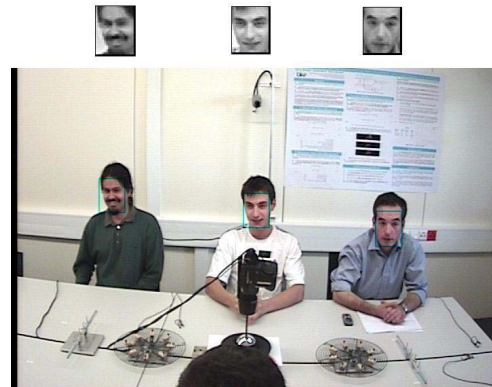

Figure 4: We cropped the face region in frontal and near frontal view based on the location of two eyes from the input image sequence (Frame 17130).

GT vs original GT are shown in Table 3.

Original GT	Neutral	Joy	Neutral	Neutral	Joy
Modified GT	Sideview	Sideview	Joy	Joy	Neutral

Table 3: Examples of modified GT vs original GT.

A total of 1209 images from the Cohn-Kanade database were used to train the SVM classifier. Since face regions in PETS dataset are around 40×50 pixels, the training images were down-sampled from the original images to 38×48 pixels. The trained classifier recognized 5 expressions: neutral, joy, angry, surprise, and others (including fear, sadness, and disgust).

Modified GT	Joy	Joy	Neutral	Neutral	Neutral
Test Results	Others	Others	Joy	Others	Joy

Table 4: Examples of failed recognition.

Our method performed well with the input real-world image sequence. To compare with Tian's work [1], we also report expression recogniton results from Frame 18000 to Frame 18190. The overall recognition rate was 91.5%, which is comparable to Tian's method. Table 4 shows some failed examples. Many frames of near frontal view were wrong classified because our training data includes only frontal view expressions. Since in our training data, most joy expression images are with opened mouth, some frames with joy expression but without opened mouth were not classified correctly.

5. Discussions and Conclusions

Although many recent advances and successes in automatic facial expression analysis have been achieved, it is difficult to apply the existing systems to real-life application. One challenge is recognizing facial expressions at low resolution, since input images in real-world environments are often low resolution.

This paper focuses on low-resolution facial expression recognition. Local Binary Patterns were introduced as novel low-cost discriminative features for low-resolution facial expression recognition. Compared to Gabor wavelets, LBP features can be derived rapidly in a single scan of raw images, whilst still retaining enough facial infomation in a compact representation. Extensive experiments on the Cohn-Kanade database demonstrated that the LBP features are effective and efficient for facial expression recognition, and crucially perform well over low resolutions. Its encouraging performance on real-world video sequences in PETS 2003 dataset illustrated its promising application in real-world environments.

There are still considerable challenges in applying automatic facial expression recognition to real-life environments. The posed expressions data studied predominantly to date are high intensity, while facial expressions in real life are often low intensity; so further efforts will be required for modeling low-intensity facial expressions. Additionally, more accurate results may be achieved if using the temporal information in video sequences.

References

[1] Y. Tian, L. Brown, A. Hampapur, S. Pankanti, A. Senior, and R. Bolle. Real world real-time automatic recognition of facial expression. In *IEEE workshop on performance evaluation of tracking and surveillance (PETS)*, Australia, March 2003.

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

[3] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36:259–275, 2003.

[4] Y. Yacoob and L. S. Davis. Recognizing human facial expression from long image sequences using optical flow. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1996.

[5] I. Essa and A. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997.

[6] M. J. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(12):1357–1362, 1999.

[7] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999.

[8] M. Pantic and L. Rothkrantz. Expert system for automatic analysis of facial expression. *Image and Vision Computing*, 18(11):881–905, 2000.

[9] Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 2001.

[10] I. Cohen, N. Sebe, A. Garg, L. Chen, and T. S. Huang. Facial expression recognition from video sequences: Temporal and static modeling. *Computer Vision and Image Understanding*, 91:160–187, 2003.

[11] Y. Zhang and Q. Ji. Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.

[12] Y. Tian, T. Kanade, and J.F. Cohn. Facial Expression Analysis in *Handbook of Face Recognition*, Springer, 2003.

[13] Z. Zhang. Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron. *International Journal of Pattern Recognition and Artificial Intelligence*, 1999.

[14] I. Buciu, C. Kotropoulos, and I. Pitas. Ica and gabor representation for facial expression recognition. In *IEEE International Conference on Image Processing (ICIP)*, 2003.

[15] M.S. Bartlett, G. Littlewort, I. Fasel, and R. Movellan. Real time face detection and facial expression recognition: Development and application to human computer interaction. In *CVPR Workshop on CVPR for HCI*, 2003.

[16] Y. Tian. Evaluation of face resolution for expression analysis. In *CVPR Workshop on Face Processing in Video*, 2004.

[17] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, , and J. Movellan. Dynamics of facial expression extracted automatically from video. In *CVPR Workshop on Face Processing in Video*, 2004.

[18] T. Ojala, M Pietikinen, and D. Harwood. A comparative study of texture measures with classification based on featured distribution. *Pattern Recognition*, 29(1):51–59, 1996.

[19] T. Ojala, M. Pietikinen, and T. Menp. Multiresolution grayscale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

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

[21] A. Hadid, M. Pietikinen, and T. Ahonen. A discriminative feature space for detecting and recognizing faces. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.

[22] T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2000.

[23] V. N. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.