Image and Vision Computing 27 (2009) 803–816

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Facial expression recognition based on Local Binary Patterns: A comprehensive study

Caifeng Shan [a,*], Shaogang Gong [b], Peter W. McOwan [b]

[a] Philips Research, High Tech Campus 36, 5656 AE Eindhoven, The Netherlands
[b] Department of Computer Science, Queen Mary, University of London, Mile End Road, London E1 4NS, UK

ARTICLE INFO

Article history:
Received 12 June 2006
Received in revised form 14 February 2008
Accepted 16 August 2008

Keywords:
Facial expression recognition
Local Binary Patterns
Support vector machine
Adaboost
Linear discriminant analysis
Linear programming

ABSTRACT

Automatic facial expression analysis is an interesting and challenging problem, and impacts important applications in many areas such as human–computer interaction and data-driven animation. Deriving an effective facial representation from original face images is a vital step for successful facial expression recognition. In this paper, we empirically evaluate facial representation based on statistical local features, Local Binary Patterns, for person-independent facial expression recognition. Different machine learning methods are systematically examined on several databases. Extensive experiments illustrate that LBP features are effective and efficient for facial expression recognition. We further formulate Boosted-LBP to extract the most discriminant LBP features, and the best recognition performance is obtained by using Support Vector Machine classifiers with Boosted-LBP features. Moreover, we investigate LBP features for low-resolution facial expression recognition, which is a critical problem but seldom addressed in the existing work. We observe in our experiments that LBP features perform stably and robustly over a useful range of low resolutions of face images, and yield promising performance in compressed low-resolution video sequences captured in real-world environments.

1. Introduction

Facial expression is one of the most powerful, natural and immediate means for human beings to communicate their emotions and intensions. Automatic facial expression analysis is an interesting and challenging problem, and impacts important applications in many areas such as human–computer interaction and data-driven animation. Due to its wide range of applications, automatic facial expression recognition has attracted much attention in recent years [1–4]. Though much progress has been made [5–24], recognizing facial expression with a high accuracy remains difficult due to the subtlety, complexity and variability of facial expressions.

Deriving an effective facial representation from original face images is a vital step for successful facial expression recognition. There are two common approaches to extract facial features: geometric feature-based methods and appearance-based methods [4]. Geometric features present the shape and locations of facial components, which are extracted to form a feature vector that represents the face geometry. Recently Valstar et al. [22,23] have demonstrated that geometric feature-based methods provide similar or better performance than appearance-based approaches in Action Unit recognition. However, the geometric feature-based methods usually requires accurate and reliable facial feature detection and tracking, which is difficult to accommodate in many-ituations. With appearance-based methods, image filters, such as Gabor wavelets, are applied to either the whole-face or specific face-regions to extract the appearance changes of the face. Due to their superior performance, the major works on appearance-based methods have focused on using Gabor-wavelet representations [25,7,8,26,19]. However, it is both time and memory intensive to convolve face images with a bank of Gabor filters to extract multi-scale and multi-orientational coefficients.

In this work, we empirically study facial representation based on Local Binary Pattern (LBP) features [27,28] for person-independent facial expression recognition. LBP features were proposed originally for texture analysis, and recently have been introduced to represent faces in facial images analysis [29–31]. The most important properties of LBP features are their tolerance against illumination changes and their computational simplicity. We examine different machine learning methods, including template matching, Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and the linear programming technique, to perform facial expression recognition using LBP features. Our study demonstrates that, compared to Gabor wavelets, LBP features can be derived very fast in a single scan through the raw image and lie in

* Corresponding author.
E-mail addresses: caifeng.shan@philips.com (C. Shan), sgg@dcs.qmul.ac.uk (S. Gong), pmco@dcs.qmul.ac.uk (P.W. McOwan).

0262-8856/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2008.08.005

low-dimensional feature space, while still retaining discriminative facial information in a compact representation. We further formulate Boosted-LBP by learning the most discriminative LBP features with AdaBoost, and the recognition performance of different classifiers are improved by using the Boosted-LBP features. We also evaluate the generalization ability of LBP features across different databases.

One limitation of the existing facial expression recognition methods is that they attempt to recognize facial expressions from data collected in a highly controlled environment given high resolution frontal faces [26]. However, in real-world applications such as smart meeting and visual surveillance, the input face images are often at low resolutions. Obviously low-resolution images in real-world environments make real-life expression recognition much more difficult. Recently Tian et al. [32,26] made a first attempt to recognize facial expressions at low resolutions. In [26], Tian studied the effects of different image resolutions for each step of automatic facial expression recognition. In this work, we investigate LBP features for low-resolution facial expression recognition. Experiments on different image resolutions show that LBP features perform stably and robustly over a useful range of low resolutions of face images. The encouraging performance on real-world compressed video sequences illustrated their promising applications in real-world environments.

This paper is an extended version of our previous work described in [33]. The main contributions of this paper are summarized as follows:

- We empirically evaluate LBP features for person-independent facial expression recognition. Different machine learning methods are exploited to classify expressions on several databases. LBP features were previously used for facial expression classification in [31], and more recently, following our work [33], Liao et al. [34] presented an extended LBP operator to extract features for facial expression recognition. However, these existing works were conducted on a very small database (JAFFE) using an individual classifier. In contrast, here we comprehensively study LBP features for facial expression recognition with different classifiers on much larger databases.
- We investigate LBP features for low-resolution facial expression recognition, a critical problem but seldom addressed in the existing work. We not only perform evaluation on different image resolutions, but also conduct experiments in real-world compressed video sequences. Compared to the previous work [32,26], LBP features provide just as good or better performance, so are very promising for real-world applications.
- We formulate Boosted-LBP by learning the most discriminative LBP histograms with AdaBoost for each expression, and the recognition performance of different classifiers are improved by using the Boosted-LBP features. We also evaluate the generalization ability of LBP features cross different databases.

The remainder of this paper is structured as follows. We present a brief review of related work in the next section. Local Binary Patterns are introduced in Section 4. Section 5 discusses facial expression recognition using LBP features with different classification techniques. We investigate low-resolution expression recognition in Section 6. Boosting LBP for expression recognition is presented in Section 7. We also evaluate across-dataset generalization in Section 8. Finally, Section 9 concludes the paper.

2. Previous work

Automatic facial expression recognition has attracted much attention from behavioral scientists since the work of Darwin in 1872 [35]. Suwa et al. [36] made the first attempt to automatically analyze facial expressions from image sequences in 1978. Much progress has been made in the last decade, and a thorough survey of the exiting work can be found in [1,2]. Here we briefly review some previous work in order to put our work in context.

2.1. Facial representation

Automatic facial expression recognition involves two vital aspects: facial representation and classifier design [4]. Facial representation is to derive a set of features from original face images to effectively represent faces. The optimal features should minimize within-class variations of expressions while maximize between-class variations. If inadequate features are used, even the best classifier could fail to achieve accurate recognition. In some existing work [5,6,14,13], optical flow analysis has been used to model muscles activities or estimate the displacements of feature points. However, flow estimates are easily disturbed by the nonrigid motion and varying lighting, and are sensitive to the inaccuracy of image registration and motion discontinuities [18]. Facial geometry analysis has been widely exploited in facial representation [9,10,16,17,24], where shapes and locations of facial components are extracted to represent the face geometry. For example, Zhang et al. [25] used the geometric positions of 34 fiducial points as facial features to represent facial images. In image sequences, the facial movements can be qualified by measuring the geometrical displacement of facial feature points between the current frame and the initial frame. Valstar et al. [22] presented AU detection by classifying features calculated from tracked fiducial facial points. Their method detects a similar amount of AUs with similar or higher recognition rates than those reported in [10,3,37]. So they argued that the facial representation based on tracked facial points is well suited for facial expression analysis. Recently they [23] further presented a fully automatic AU detection system that can automatically localize facial points in the first frame and recognize AU temporal segments using a subset of most informative spatio-temporal features selected by AdaBoost. However, the geometric feature-based representation commonly requires accurate and reliable facial feature detection and tracking, which is difficult to accommodate in many situations. In [18], motions of facial features are measured by simultaneously using an active Infra-Red illumination and Kalman filtering to deal with large variations of head motion.

Another kind of method to represent faces is to model the appearance changes of faces. Holistic spatial analysis including Principal Component Analysis (PCA) [38], Linear Discriminant Analysis (LDA) [39], Independent Component Analysis (ICA) [40] and Gabor wavelet analysis [7] have been applied to either the whole-face or specific face regions to extract the facial appearance changes. Donato et al. [8] explored different techniques to represent face images for facial action recognition, which include PCA, ICA, Local Feature Analysis (LFA), LDA and local schemes such as Gabor-wavelet representation and local principal components. Best performances were obtained using Gabor-wavelet representation and ICA. Due to their superior performance, Gabor-wavelet representations have been widely adopted in face image analysis [25,7,26,19]. However, the computation of Gabor-wavelet representations is both time and memory intensive, for example, in [19], the Gabor-wavelet representation derived from each 48×48 face image has the high dimensionality of $O(10^5)$. Recently Local Binary Patterns have been introduced as effective appearance features for facial image analysis [31,29,30]. We [33] compared LBP features with Gabor features for facial expression recognition, and studied their performance over a range of image resolutions. In [41], we further presented facial expression manifold learning in the LBP feature space. More recently, Liao et al. [34] introduced

an improved LBP operator to extract features in both intensity and gradient maps for facial expression recognition, and also tested their methods on facial images of reduced resolutions. However, their experiment was carried out on a very small database (213 images from 10 subjects). In this work, we comprehensively study LBP features for facial expression recognition on several databases.

2.2. Facial expression recognition

Different techniques have been proposed to classify facial expressions, such as Neural Network [42,25,26], Support Vector Machine (SVM) [19], Bayesian Network (BN)[11] and rule-based classifiers [9,17,24]. In Lyons et al.' work [7], the principle components of the feature vectors from training images were analyzed by LDA to form discriminant vectors, and facial image classification was performed by projecting the input vector of a testing image along the discriminant vectors. Cohen et al. compared different Bayes classifiers [11], and Gaussian Tree-Augmented-Naive (TAN) Bayes classifiers performed best. Bartlett et al. [19] performed systematic comparison of different techniques including AdaBoost, SVM and LDA for facial expression recognition, and best results were obtained by selecting a subset of Gabor filters using AdaBoost and then training SVM on the outputs of the selected filters. Pantic and Rothkrantz adopted rule-based reasoning to recognize action units and their combination [17].

To exploit the temporal behaviors of facial expressions, different techniques were presented for facial expression recognition in image sequences. There have been several attempts to track and recognize facial expressions over time based on optical flow analysis [5,6]. Tian et al. [10] presented a Neural Network based approach to recognize facial action units in image sequences. Hidden Markov Models (HMMs) have been widely used to model the temporal behaviors of facial expressions from image sequences [11,13]. Cohen et al. [11] proposed a multi-level HMM classifier, which allows not only to perform expression classification on a video segment, bust also to automatically segment a long video sequence to the different expressions segments without resorting to heuristic methods of segmentation. But HMMs can not deal with dependencies in observation. Dynamic Bayesian Networks (DBNs) recently were exploited for sequence-based expression recognition [16,14,18]. Kaliouby and Robinson [16] proposed a system for inferring complex mental states from videos of facial expressions and head gestures, where a multi-level DBN classifier was used to model complex mental states as a number of interacting facial and head displays. Zhang and Ji [18] explored the use of multisensory information fusion technique with DBNs for modeling and understanding the temporal behaviors of facial expressions in image sequences. Chang et al. proposed a probabilistic video-based facial expression recognition method based on manifolds [15]. Lee and Elgammal [21] recently introduced a framework to learn decomposable generative models for dynamic appearance of facial expressions where facial motion is constrained to one dimensional closed manifolds. The learned model can generate different dynamic facial appearances for different people and for different expressions, so enabling simultaneous recognition of faces and facial expressions.

3. Facial expression data

Facial expressions can be described at different levels [4]. A widely used description is Facial Action Coding System (FACS) [43], which is a human-observer-based system developed to capture subtle changes in facial expressions. With FACS, facial expressions are decomposed into one or more Action Units (AUs). AU recognition or detection has attracted much attention recently [8,10,18,23]. Meanwhile, psychophysical studies indicate that basic emotions have corresponding universal facial expressions across all cultures [44]. This is reflected by most current facial expression recognition systems [7,11–13,19] that attempt to recognize a set of prototypic emotional expressions including disgust, fear, joy, surprise, sadness and anger. Therefore, in this work, we also focus on prototypic expression recognition. We consider both 6-class prototypic expression recognition and 7-class expression recognition by including the neutral expression.

We mainly conducted experiments on the Cohn–Kanade database [45], one of the most comprehensive database in the current facial-expression-research community. The database consists of 100 university students aged from 18 to 30 years, of which 65% were female, 15% were African-American and 3% were Asian or Latino. Subjects were instructed to perform a series of 23 facial displays, six of which were based on description of prototypic emotions. Image sequences from neutral to target display were digitized into 640×490 pixel arrays with 8-bit precision for gray-scale values. Fig. 1 shows some sample images from the Cohn–Kanade database.

For our experiments, we selected 320 image sequences from the database. The only selection criterion was that a sequence could be labeled as one of the six basic emotions. The sequences come from 96 subjects, with 1–6 emotions per subject. For each sequence, the neutral face and three peak frames were used for prototypic expression recognition, resulting in 1280 images (108 Anger, 120 Disgust, 99 Fear, 282 Joy, 126 Sadness, 225 Surprise and 320 Neutral). To evaluate the generalization performance to novel subjects, we adopted a 10-fold cross-validation testing scheme in our experiments. More precisely, we partitioned the dataset randomly into ten groups of roughly equal numbers of subjects. Nine groups were used as the training data to train classifiers, while the remaining group was used as the test data. The above process was repeated ten times for each group in turn to be omitted from the training process. We reported the average recognition results on the test sets.

Following Tian [26], we normalized the faces to a fixed distance between the two eyes. We manually labeled the eyes location, to evaluate LBP features in the condition of no face registration errors. Automatic face registration can be achieved by face detection [46] and eye localization [26,47], which will be addressed in our future work. Facial images of 110×150 pixels were cropped from original frames based on the two eyes location. No further registration such as alignment of mouth [25] was performed in our algorithms. As the faces in the database are frontal view, we did not consider head pose changes. For realistic sequences with head pose variation, head pose estimation [26] can be adopted to detect front or near front view. Illumination changes exist in the database, but there was no attempt made to remove illumination changes [26] in our experiments, due to LBP's gray-scale invariance. Fig. 2 shows an example of the original face image and the cropped image.

4. Local Binary Patterns (LBP)

The original LBP operator was introduced by Ojala et al. [27], and was proved a powerful means of texture description. The operator labels the pixels of an image by thresholding a 3×3 neighborhood of each pixel with the center value and considering the results as a binary number (see Fig. 3 for an illustration), and the 256-bin histogram of the LBP labels computed over a region is used as a texture descriptor. The derived binary numbers (called Local Binary Patterns or LBP codes) codify local primitives including different types of curved edges, spots, flat areas, etc (as shown in Fig. 4), so each LBP code can be regarded as a micro-texton [30].

The limitation of the basic LBP operator is its small 3×3 neighborhood which can not capture dominant features with large scale

Fig. 1. The sample face expression images from the Cohn–Kanade database.

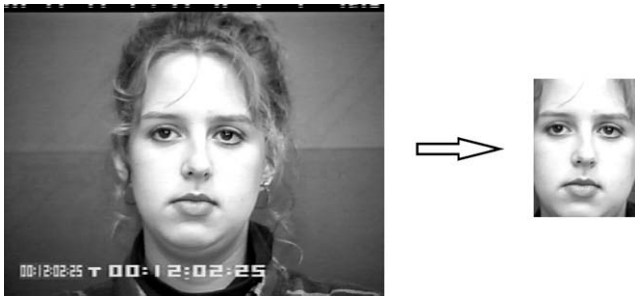

Fig. 2. The original face image and the cropped image.

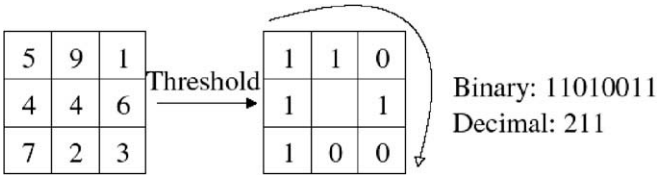

Fig. 3. The basic LBP operator [29].

structures. Hence the operator later was extended to use neighborhood of different sizes [28]. Using circular neighborhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighborhood. See Fig. 5 for 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 circularly symmetric neighbor 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 neighbor set. It has been shown that certain bins contain more information than others [28]. Therefore, it is possible to use only a subset of the 2^P Local Binary Patterns to describe the texture of images. Ojala et al. [28] called these fundamental patterns as uniform patterns. A Local Binary Pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or 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 the (8,1) neighborhood and for about 70% in the (16,2) neighborhood in texture images [28]. Accumulating the patterns which have more than 2 transitions into a single bin yields an LBP operator, denoted $LBP^{u2}_{P,R}$, with less than 2^P bins. For example, the number of labels for a neighborhood of 8 pixels is 256 for the standard LBP but 59 for LBP^{u2}.

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

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

where n is the number of different labels produced by the LBP operator and

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

This LBP histogram contains information about the distribution of the local micro-patterns, such as edges, spots and flat areas, over the whole image, so can be used to statistically describe image characteristics.

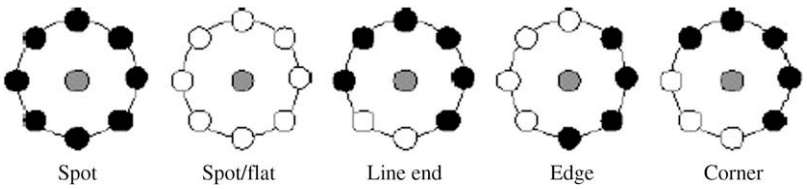

Fig. 4. Examples of texture primitives which can be detected by LBP (white circles represent ones and black circles zeros) [30].

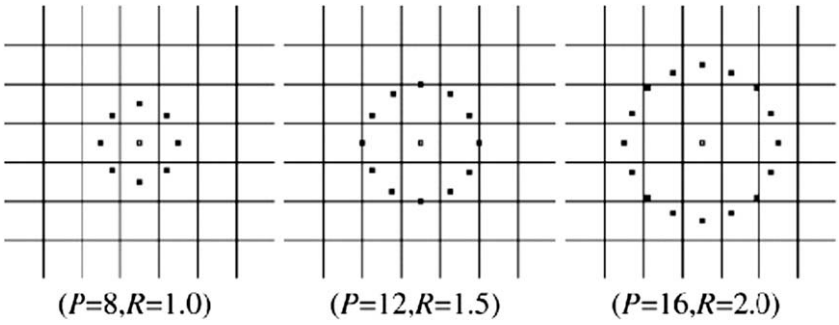

Fig. 5. Three examples of the extended LBP [28]: the circular $(8,1)$ neighborhood, the circular $(12,1.5)$ neighborhood, and the circular $(16,2)$ neighborhood, respectively.

Face images can be seen as a composition of micro-patterns which can be effectively described by the LBP histograms. Therefore, it is intuitive to use LBP features to represent face images [29–31]. 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 information of faces, face images were equally divided into small regions $R_0, R_1, \ldots, R_m$ to extract LBP histograms (as shown in Fig. 6). The LBP features extracted from each sub-region are concatenated into a single, spatially enhanced feature histogram defined as

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

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

The extracted feature histogram represents the local texture and global shape of face images. Some parameters can be optimized for better feature extraction. One is the LBP operator, and the other is the number of regions divided. Following the setting in [29], we selected the 59-bin $LBP^{u2}_{8,2}$ operator, and divided the 110×150 pixels face images into 18×21 pixels regions, giving a good trade-off between recognition performance and feature vector length. Thus face images were divided into $42(6 \times 7)$ regions as shown in Fig 7, and represented by the LBP histograms with the length of $2478(59 \times 42)$.

5. Facial expression recognition using LBP

In this section, we perform person-independent facial expression recognition using LBP features. Different machine learning techniques, including template matching, Support Vector Machines, Linear Discriminant Analysis and the linear programming technique, are examined to recognize expressions.

5.1. Template matching

Template matching was used in [29] to perform face recognition using the LBP-based facial representation: a template is

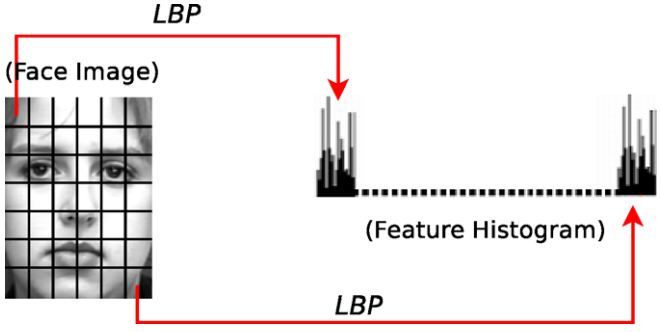

Fig. 6. A face image is divided into small regions from which LBP histograms are extracted and concatenated into a single, spatially enhanced feature histogram.

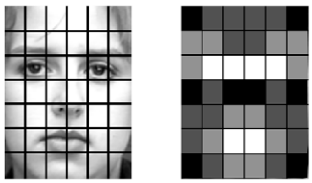

Fig. 7. (Left) A face image divided into 6×7 sub-region. (Right) The weights set for weighted dissimilarity measure. Black squares indicate weight 0.0, dark gray 1.0, light gray 2.0 and white 4.0.

formed for each class of face images, then a nearest-neighbor classifier is used to match the input image with the closest template. Here we first adopted template matching to classify facial expressions for its simplicity. In training, the histograms of expression images in a given class were averaged to generate a template for this class.

Following [29], we also selected the Chi square statistic (χ^2) as the dissimilarity measure for histograms:

$$\chi^2(\mathbf{S}, \mathbf{M}) = \sum_i \frac{(S_i - M_i)^2}{S_i + M_i} \tag{4}$$

where **S** and **M** are two LBP histograms. It is observed that some local facial regions contain more useful information for expression classification than others. For example, facial features contributing to facial expressions mainly lie in regions such as eye and mouth regions. Therefore, a weight can be set for each sub-region based on its importance. The particular weight set we adopted was shown in Fig. 7, which was designed empirically based on the observation. The weighted χ^2 statistic is then given as

$$\chi^2_w(\mathbf{S}, \mathbf{M}) = \sum_{i,j} w_j \frac{(S_{i,j} - M_{i,j})^2}{S_{i,j} + M_{i,j}} \tag{5}$$

where **S** and **M** are two LBP histograms, and w_j is the weight for region j.

The template matching achieved the generalization performance of 79.1% for the 7-class task and 84.5% for the 6-class task. We compared the results with that reported in [11], where Cohen et al. adopted Bayesian network classifiers to classify 7-class emotional expressions based on the tracked geometric facial features (eyebrows, eyelids and mouth). They carried out 5-fold cross-validation on a subset of 53 subjects from the Cohn–Kanade database, and obtained the best performance of 73.2% by using Tree-Augmented-Naive Bayes (TAN) classifiers. Although we cannot make a direct comparison due to different experiment setups and preprocessing procedures, comparison in Table 1 indicates that our simple template matching using LBP features provides slightly better overall performance. The confusion matrix of 7-class recognition is shown in Table 2. We can observe that Joy and Surprise can be recognized with high accuracy (around 90–92%), but Anger and Fear are easily confused with others.

Table 1
Comparisons between the geometric features based TAN [11] and our LBP-based template matching

Methods (feature + classifier)	7-Class recognition (%)	6-Class recognition (%)
LBP + template matching	79.1 ± 4.6	84.5 ± 5.2
Geometric features + TAN [11]	73.2	–

Table 2
Confusion matrix of 7-class facial expression recognition using template matching with LBP features

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sadness (%)	Surprise (%)	Neutral (%)
Anger	**58.7**	5.5	0	0	26.7	0	9.1
Disgust	3.3	**85.0**	2.5	0	2.5	0	6.7
Fear	1.0	0	**61.7**	24.0	10.3	0	3.0
Joy	0	0	6.0	**90.4**	0	0	3.6
Sadness	4.9	0	0	0	**72.4**	1.7	21.0
Surprise	0	0	1.3	0	2.7	**92.4**	3.6
Neutral	2.0	0.8	0.4	0.8	25.7	0	**70.3**

5.2. Support Vector Machine (SVM)

A previous successful technique to facial expression classification is Support Vector Machine (SVM) [48,19,22,23], so we adopted SVM as alternative classifiers for expression recognition. As a powerful machine learning technique for data classification, SVM [49] performs an implicit mapping of data into a higher (maybe infinite) dimensional feature space, and then finds a linear separating hyperplane with the maximal margin to separate data in this higher dimensional space.

Given a training set of labeled examples $\{(x_i, y_i), i = 1, \ldots, l\}$ where $x_i \in R^n$ and $y_i \in \{1, -1\}$, a new test example x is classified by the following function:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b\right) \quad (6)$$

where α_i are Lagrange multipliers of a dual optimization problem that describe the separating hyperplane, $K(\cdot,\cdot)$ is a kernel function, and b is the threshold parameter of the hyperplane. The training sample x_i with $\alpha_i > 0$ is called *support vectors*, and SVM finds the hyperplane that maximizes the distance between the support vectors and the hyperplane. Given a non-linear mapping Φ that embeds the input data into the high dimensional space, kernels have the form of $K(x_i, x_j) = \langle \Phi(x_i) \cdot \Phi(x_j) \rangle$. SVM allows domain-specific selection of the kernel function. Though new kernels are being proposed, the most frequently used kernel functions are the linear, polynomial, and Radial Basis Function (RBF) kernels.

SVM makes binary decisions, so the multi-class classification here is accomplished by using the one-against-rest technique, which trains binary classifiers to discriminate one expression from all others, and outputs the class with the largest output of binary classification. With regard to the parameter selection of SVM, as suggested in [50], we carried out grid-search on the hyper-parameters in the 10-fold cross-validation. The parameter setting producing best cross-validation accuracy was picked. We used the SVM implementation in the public available machine learning library SPIDER[1] in our experiments. The generalization performances achieved with different kernels are shown in Table 3, where the degree of the polynomial kernel is 1, and the standard deviation for the RBF kernel is 2^{13} for 7-class recognition and 2^{11} for 6-class recognition. The confusion matrices of 6-class and 7-class recognition with the RBF kernel are shown in Tables 4 and 5. It is observed that, Disgust, Joy, Surprise and Neutral can be recognized with high accuracy (90–98%), while the recognition rates for Fear and Sadness are much lower (68–69%). Compared to the recognition results of template matching in Table 2, the recognition performance for every expression is increased except Fear. For the 6-class problem, the number of support vectors of the linear/polynomial SVMs were 18–29% of the total number of training samples, while the RBF SVMs employed 18–31%. For the 7-class problem, the linear/polynomial SVMs employed 15–30%, while the RBF SVMs employed 16–35%.

Table 3
Recognition performance of LBP-based SVM with different kernels

	6-Class recognition (%)	7-Class recognition (%)
SVM (linear)	91.5 ± 3.1	88.1 ± 3.8
SVM (polynomial)	91.5 ± 3.1	88.1 ± 3.8
SVM (RBF)	92.6 ± 2.9	88.9 ± 3.5

Table 4
Confusion matrix of 6-class facial expression recognition using SVM (RBF)

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sadness (%)	Surprise (%)
Anger	**89.7**	2.7	0	0	7.6	0
Disgust	0	**97.5**	2.5	0	0	0
Fear	0	2.0	**73.0**	22.0	3.0	0
Joy	0	0.4	0.7	**97.9**	1.0	0
Sadness	10.3	0	0.8	0.8	**83.5**	4.6
Surprise	0	0	1.3	0	0	**98.7**

Table 5
Confusion matrix of 7-class facial expression recognition using SVM (RBF)

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sadness (%)	Surprise (%)	Neutral (%)
Anger	**85.0**	2.7	0	0	4.8	0	7.5
Disgust	0	**97.5**	2.5	0	0	0	0
Fear	0	2.0	**68.0**	22.0	1.0	0	7.0
Joy	0	0	0.7	**94.7**	1.1	0	3.5
Sadness	8.6	0	0	0	**69.5**	2.3	19.6
Surprise	0	0	1.3	0	0	**98.2**	0.5
Neutral	1.6	0.4	0	1.6	6.0	0.4	**90.0**

We further compare LBP features with Gabor-wavelet features for facial expression recognition using SVMs. Following Bartlett et al. [48,19], we converted images into a Gabor magnitude representation using a bank of Gabor filters at 8 orientations and 5 spatial frequencies (9:36 pixels per cycle at 1/2 octave steps[2]). To reduce the length of the feature vector, the outputs of the 40 Gabor filters were downsampled by a factor of 16 [8], so the dimensionality of the Gabor feature vector is $42,650 (40 \times 110/4 \times 150/4)$. We report the generalization performance of Gabor-wavelet features in Table 6.

Bartlett et al. [48,19] recently conducted similar experiments using the Gabor-wavelet representation with SVMs on the Cohn–Kanade database. They selected 313 image sequences from the database, which came from 90 subjects, with 1–6 emotions per subject. The facial images were converted into a Gabor magnitude representation using a bank of 40 Gabor filters. They [48] divided the subjects randomly into ten groups of roughly equal size and did "leave one group out" cross-validation. SVMs with linear, polynomial and RBF kernels were used to classify 7-class expressions. Linear and RBF kernels performed best, achieving recognition rates of 84.8% and 86.9%, respectively. We also include the recognition

[1] http://www.kyb.tuebingen.mpg.de/bs/people/spider/index.html.

[2] i.e., 9, $9\sqrt{2}$, 18, $18\sqrt{2}$, 36 pixels per cycle, so the frequencies used ≈12, $6\sqrt{2}$, 6, $3\sqrt{2}$, 3 cycles/image-width.

Table 6
Comparisons between LBP features with Gabor-filter features for facial expression recognition using SVMs.

	6-Class		7-Class		
	LBP (%)	Gabor (%)	LBP (%)	Gabor (%)	Gabor [48] (%)
SVM (linear)	91.5 ± 3.1	89.4 ± 3.0	88.1 ± 3.8	86.6 ± 4.1	84.8
SVM (polynomial)	91.5 ± 3.1	89.4 ± 3.0	88.1 ± 3.8	86.6 ± 4.1	Worse than RBF/linear
SVM (RBF)	92.6 ± 2.9	89.8 ± 3.1	88.9 ± 3.5	86.8 ± 3.6	86.9

results they reported in Table 6. In their more recent paper [19], they reported 88.0% (Linear) and 89.1% (RBF) in Leave-one-subject-out experiments.

Comparisons summarized in Table 6 show that the LBP-based SVMs perform slightly better than the Gabor-wavelet based SVMs. More crucially though, the advantage of LBP features lies at very fast feature extraction. We compare the time and memory costs of feature extraction process (Matlab implementation) between LBP features with Gabor-wavelet features in Table 7, where the Gabor-filter convolutions were calculated in spatial domain. It is observed that LBP features bring significant speed benefit, and, compared to the high dimensionality of the Gabor-wavelet features, LBP features lie in a much lower dimensional space.

5.3. Linear Discriminant Analysis (LDA)

Facial deformations lie intrinsically on much lower dimensional subspaces. Therefore, subspace analysis has been widely exploited to discover subspaces for face image analysis [38,39,42,7]. LDA [39] is a supervised subspace learning technique, and has been previously applied to facial expression recognition [7]. Here we further adopted LDA to recognize expressions using LBP features. LDA searches for the projection axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other.

Given multi-dimensional data samples $x_1, x_2, \ldots, x_m$ in R^n that belong to c classes, LDA find a transformation matrix W that maps these m points to $y_1, y_2, \ldots, y_m$ in $R^l (l \leqslant c)$, where $y_i = W^{\mathrm{T}} x_i$. The objective function of LDA is as follows:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^{\mathrm{T}} S_B \mathbf{w}}{\mathbf{w}^{\mathrm{T}} S_W \mathbf{w}} \tag{7}$$

$$S_B = \sum_{i=1}^{c} n_i (\mathbf{m}^{(i)} - \mathbf{m})(\mathbf{m}^{(i)} - \mathbf{m})^{\mathrm{T}} \tag{8}$$

$$S_W = \sum_{i=1}^{c} \left(\sum_{j=1}^{n_i} (x_j^{(i)} - \mathbf{m}^{(i)})(x_j^{(i)} - \mathbf{m}^{(i)})^{\mathrm{T}} \right) \tag{9}$$

where $\mathbf{m}$ is the mean of all the samples, n_i is the number of samples in the ith class, $\mathbf{m}^{(i)}$ is the average vector of the ith class, $x_j^{(i)}$ is the jth sample in the ith class, S_B is between-class scatter matrix, and S_W is within-class scatter matrix. In practice, the dimension of the feature space (n) is often much larger than the number of samples in a training set (m). So the matrix S_W is singular. To overcome this problem, usually the dataset is first projected into a lower dimensional PCA space.

Table 7
Time and memory costs for extracting LBP features and Gabor-filter features

	LBP	Gabor	Gabor [48]
Memory (feature dimension)	2478	42,650	92,160
Time (feature extraction time)	0.03 s	30 s	–

In each trial of our 10-fold cross-validation experiments, the training data was first projected into a PCA subspace (98% of information was kept according to the reconstruction error, and the resulting number of eigenvectors ranges 358–378 for the 6-class problem and 405–431 for the 7-class problem), then the LDA transformation matrix was trained in the PCA subspace, where the dimension that LDA kept was $c - 1$. For facial expression recognition, we adopted a Nearest-Neighbor (NN) classifier for its simplicity. The Euclidean metric was used as the distance measure. The generalization performance LDA + NN achieved is 73.4% for 7-class recognition and 79.2% for 6-class recognition. We also adopted SVM (linear) to perform recognition in the PCA subspace, i.e., the input of SVM is not the original LBP features but their PCA projections. We compare LDA + NN with SVM (linear) in Table 8, and it is observed that the performance of LDA + NN is much lower than that of SVMs.

5.4. Linear programming

Feng et al. [51] recently presented an approach for facial expression recognition that uses LBP features with a linear programming technique, and demonstrated its effectiveness on a small database (JAFFE). In [52], the linear programming technique was adopted to perform simultaneous feature selection and classifier training for facial expression recognition. Here we also examine the linear programming technique for facial expression recognition using LBP features.

Given two sets of data samples $\mathscr{A}$ and $\mathscr{B}$ in R^n, we seek a linear function such that $f(x) > 0$ if $x \in \mathscr{A}$, and $f(x) \leqslant 0$ if $x \in \mathscr{B}$. This function is given by $f(x) = w^{\mathrm{T}} x - \gamma$, and determine a plane $w^{\mathrm{T}} x = \gamma$ with normal $w \in R^n$ that separate $\mathscr{A}$ from $\mathscr{B}$. Let the set of m samples in $\mathscr{A}$ be represented by a matrix $A \in R^{m \times n}$ and the set of k samples in $\mathscr{B}$ be represented by a matrix $B \in R^{k \times n}$. After normalization, we want to satisfy

$$Aw \geqslant e\gamma + e, \quad Bw \leqslant e\gamma - e \tag{10}$$

where e is a vector of all 1s with appropriate dimension. Practically, because of the overlap between the two classes, one has to minimize some norm of the average error in Eq. (10) [52]:

$$\min_{w,\gamma} f(w,\gamma) = \min_{w,\gamma} \frac{1}{m} \|(-Aw + e\gamma + e)_+\|_1 + \frac{1}{k} \|(Bw - e\gamma + e)_+\|_1 \tag{11}$$

where x_+ denotes the vector with components satisfying $(x_+)_i = max\{x_i, 0\}, i = 1, \ldots, n$, and $\|\cdot\|_1$ denotes the 1-norm. Eq. (11) can be modeled as a so-called robust linear programming problem [52]:

$$\min_{w,\gamma,y,z} \frac{e^{\mathrm{T}} y}{m} + \frac{e^{\mathrm{T}} z}{k} \tag{12}$$

$$\text{subject to} \begin{cases} -Aw + e\gamma + e \leqslant y, \\ Bw - e\gamma + e \leqslant z, \\ y \geqslant 0, \quad z \geqslant 0 \end{cases}$$

which minimizes the average sum of misclassification errors. We use Eq. (12) to solve the classification problem.

Following Feng et al. [51], multi-class facial expression recognition was decomposed into one-to-one pairs of binary classification,

Table 8
Comparison between LDA + NN and SVM (linear) for facial expression recognition using LBP features

	7-Class recognition (%)	6-Class recognition (%)
LDA + NN	73.4 ± 5.6	79.2 ± 7.2
SVM (linear)	80.2 ± 4.9	87.7 ± 4.7

where each binary classifier was produced by the linear programming technique. Binary classifiers were combined with a voting scheme to output the final recognition result. To reduced the length of the LBP feature vector, we also discarded the dimensions whose occurrence frequency is lower than a threshold [51]. The threshold of 5 was adopted in our experiments.

In our 10-fold cross-validation experiments, the linear programming technique produces the generalization performance of 82.3% for 7-class recognition and 89.6% for 6-class recognition. We compare its performance with that of SVM (linear) in Table 9, where the input of SVM (linear) is also the feature vectors with dimensions discarded. It is observed that the linear programming technique produces the slight inferior performance to SVM (linear).

Table 9
Comparison between the linear programming technique and SVM (linear) for facial expression recognition

	7-Class recognition (%)	6-Class recognition (%)
Linear programming	82.3 ± 3.5	89.6 ± 3.6
SVM (linear)	86.0 ± 3.3	90.4 ± 3.9

6. Low-resolution facial expression recognition

In real-world environments such as smart meeting and visual surveillance, only low-resolution video input is available. Fig. 8 shows a real-world image recorded in a smart meeting scenario. How to derive a discriminative facial representation from low-resolution images is a critical problem for real-world applications. In this section, we investigate LBP features for low-resolution facial expression recognition. We first evaluated LBP features on different image resolutions, then performed experiments on real-world compressed low-resolution video sequences.

6.1. Evaluation on different resolutions

As shown in Table 10, 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). For example, face images of 14×19 pixels were divided into $12(3 \times 4)$ regions of 10×10 pixels: the overlap between adjacent regions is 8 pixels (along the side of 14 pixels) or/and 7 pixels (along the side of 19 pixels). We adopted the 4-neighborhood LBP operator $LBP_{4,1}$ for each subregion.

To compare with Tian's work [26], we conducted experiments on 6-class basic expression recognition using SVM with RBF kernel. We report the recognition results in Table 10, where the standard deviation of RBF kernels were 2^{11}, 2^9, 2^7, 2^8, 2^6 and 2^8, respectively. Besides LBP features, We also carried out experiments with the Gabor-magnitude representation by convolving images with a bank of 40 Gabor filters at 8 orientations and 5 spatial frequencies. The generalization performances of the Gabor-wavelet representation are also shown in Table 10.

In Tian's experiments [26], 375 image sequences were selected from the Cohn–Kanade database for 6-class expression classification. Tian extracted two types of facial features: geometric features and appearance features. Geometric features were derived by feature tracking [10] and feature detection [32], respectively. For appearance features, a bank of 40 Gabor filters were applied to the difference images to extract facial appearance changes, where the difference images were obtained by subtracting a neural expression for each image. A three-layer Neural Network was adopted to recognize expressions. Recognition results of Tian's methods are summarized in Table 10.[3] However, we cannot make direct comparative analysis between Tian's results with ours because of different experimental setups, pre-processing procedures and classifiers.

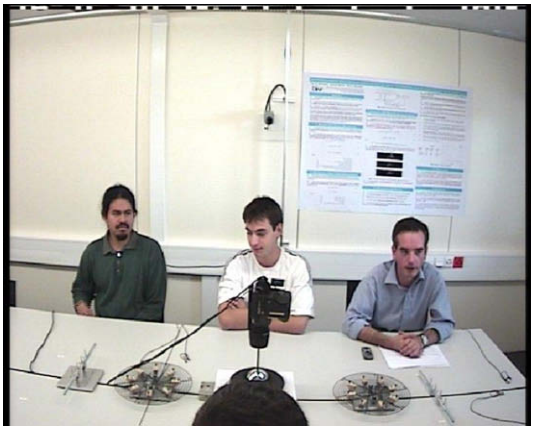

Fig. 8. An example of low-resolution facial expressions recorded in real-world environments (from PETS 2003 dataset).

We can draw the following conclusions from the experimental results shown in Table 10: (1) Geometric features are not available for lower resolution, while appearance features such as Gabor wavelets and LBP features can be extracted on different resolutions. It is difficult to detect or track facial components such as mouth, eyes, brows and nose in lower resolution images, so geometric features are not reliable in low-resolution images. On the contrary, appearance features present the appearance changes of faces such as wrinkles and furrows, and are available even in lower resolutions. (2) The presented LBP features perform slightly better than the Gabor-wavelet representation on low-resolution expression recognition. Recently Liao et al. [34] also compared LBP feature with Gabor-filter features on the JAFFE database, and their experiments demonstrated that LBP features provide better performance for low-resolution face images, which reenforces our finding (3) The LBP features perform robustly and stably over a useful range of low resolutions. This reenforce the superiority of LBP features in face detection and recognition in low-resolution images reported in [30]. So LBP features are very promising for real-world applications where low-resolution video input is only available.

6.2. Evaluation on real-world video sequences (PETS)

We further conducted experiments on compressed low-resolution image sequences recorded in a real environment. We used the smart meeting dataset in the PETS 2003 evaluation datasets.[4] 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 sits down. Then each person looks at the other people with different expressions. Fig. 8 shows an example frame in the video sequence. Three facial expressions, neutral, anger and joy, are available in the dataset.

[3] In [26], the different resolutions of the head region are 144×192, 72×96, 36×48, 18×24 pixels, which are comparable to the resolutions of the face region 110×150, 55×75, 27×37, 14×17 pixels in our experiments.

[4] http://www.cvg.cs.rdg.ac.uk/PETS-ICVS/pets-icvs-db.html.

Table 10
Recognition performance (%) in low-resolution images with different methods

	 110 × 150	 55 × 75	 36 × 48	 27 × 37	 18 × 24	 14 × 19
LBP	92.6 ± 2.9	89.9 ± 3.1	87.3 ± 3.4	84.3 ± 4.1	79.6 ± 4.7	76.9 ± 5.0
Gabor	89.8 ± 3.1	89.2 ± 3.0	86.4 ± 3.3	83.0 ± 4.3	78.2 ± 4.5	75.1 ± 5.1
Gabor [26]	92.2	91.6	–	77.6	–	68.2
Feature tracking [26]	91.8	91.6	–	N/A	–	N/A
Feature detection [26]	73.8	72.9	–	61.3		N/A

"–" indicates that the image resolution was not examined in [26], and "N/A" indicates that the image resolution was studied in [26], but no recognition result was obtained.

The real-world video sequence contains the full range of head motion. In Tian's previous work [32], the head pose was first estimated based on the detected head, and then for frontal and near frontal views of the face, the facial features were extracted to perform facial expression recognition. Since our focus was investigating the validity of LBP features 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 recognition on the cropped face images. Fig. 9 shows face regions cropped in one frame.

It is very difficult, even for human beings, to recognize facial expressions at low resolution. Following Tian et al. [32], experiments were conducted on showing some frames of expression at low resolution 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) for some frames based on human observations. Examples of modified GT vs original GT are shown in Table 11.

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 five expressions: neutral, joy, angry, surprise and others (including fear, sadness and disgust).

Our method performed well with the input real-world image sequence. The overall recognition rate on frames from 18,000 to 18,190 was 91.5%, which is comparable to results reported in Tian's work [32]. Table 12 shows some failed examples. We observe that some frames of near frontal view were incorrectly classified because our training data includes only frontal view expressions. Additionally, as the training images are captured when subjects exaggeratedly pose their facial expressions, while the test images are natural facial expressions without any deliberate exaggerated posing, this difference in data also brings some classification errors.

Fig. 9. We cropped the face region in frontal and near frontal view based on the location of two eyes from the input image sequence (frame 17,130).

Table 11
Examples of modified GT vs original GT

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

Table 12
Examples of failed recognition

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

7. Boosting LBP for facial expression recognition

The above experiments clearly demonstrate that the LBP features are effective for facial expression recognition, and performed just as well or better than reported existing techniques but with a significant low-computation advantage. In the above investigation, face images are equally divided into small sub-regions from which LBP histograms are extracted and concatenated into a single feature vector. However, apparently the extracted LBP features depend on the divided sub-regions, so this LBP feature extraction scheme suffers from fixed sub-region size and positions. By shifting and scaling a sub-window over face images, many more sub-regions can be obtained, bringing many more LBP histograms, which yield a more complete description of face images. To minimize a very large number of LBP histograms necessarily introduced by shifting and scaling a sub-window, boosting learning [53] can be used to learn the most effective LBP histograms that containing much discriminative information. In [54], Zhang et al. presented an approach for face recognition by boosting LBP-based classifiers, where the distance between corresponding LBP histograms of two face images is used as a discriminative feature, and AdaBoost was used to learn a few of most efficient features. In our previous work [55], we presented a conditional mutual information base boosting scheme to select the most discriminative LBP histograms for facial expression recognition. We observed that AdaBoost performs better than the conditional mutual information based boosting when using several tens of weak classifiers. Therefore, in this section, we learn the most discriminative LBP histograms using AdaBoost for better facial representation.

AdaBoost methods [56,53] provide a simple yet effective approach for stagewise learning of a nonlinear classification function. AdaBoost learns a small number of weak classifiers whose performance can be just better than random guessing, and boosts them iteratively into a strong classifier of higher accuracy. The process of AdaBoost maintains a distribution on the training samples. At each iteration, a weak classifier which minimizes the weighted error rate is selected, and the distribution is updated to increase the weights of the misclassified samples and reduce the importance of the others. AdaBoost has been successfully used in many problems such as face detection [46].

As each LBP histogram is calculated from a sub-region, AdaBoost is actually used to find the sub-regions that contain more discriminative information for facial expression classification in term of the LBP histogram. On selecting a weak classifier for AdaBoost, we adopted the histogram-based template matching. For each sub-region, the LBP histograms in a given class are averaged to generate a template for this class. The trained weak classifier matches the input histogram with the closest template, and outputs the corresponding class label. The Chi square statistic (χ^2) was used as the dissimilarity measure for histograms (Eq. (4)). As the traditional AdaBoost works on two-class problems, the multi-class problem here is accomplished by using the one-against-rest technique, which trains AdaBoost between one expression with all others. For each AdaBoost learner, the images of one expression were positive samples, while the images of all other expressions were negative samples.

By shifting and scaling a sub-window, 16,640 sub-regions, i.e., 16,640 LBP histograms, in total were extracted from each face image. The sub-window was shifted in the whole image with the shifting step of 4 pixels, while its size was scaled between 10×10 pixels and 25×20 pixels with the scaling step of 5 pixels. AdaBoost was used to learn a small subset (in tens) of effective LBP histograms. we plot in Fig. 10 the spatial localization of the 50 sub-regions (i.e., the centers of the sub-regions) that corresponded by the top 50 LBP histograms selected by AdaBoost for each expression. It is observed that different expressions have different key discriminant LBP features, and the discriminant features are mainly distributed in the eye and mouth regions.

We performed facial expression recognition using the strong classifiers boosted by AdaBoost, and outputs the class with the largest positive output of binary classifiers. In our experiments, AdaBoost training continued until the classifier output distribution for the positive and negative samples were completely separated, so the number of LBP histograms selected for each expression was not pre-defined, but automatically decided by the AdaBoost learner itself. In the 10-fold experiments, the number of selected LBP histogram ranges 49–52 for 6-class expressions and 65–70 for 7-class expressions. For example, Fig. 11 displays the selected sub-regions (LBP histograms) for each basic expression in one trial of the 10-fold cross-validation. We can observe that the selected sub-regions have variable sizes and positions. Moreover, while the weights of sub-regions in the template matching in Section 5.1 were chosen empirically, the weights in boosted classifiers were learned by AdaBoost. The generalization performance of the boosted classifiers is 84.6% for 7-class recognition and 89.8% for 6-class recognition, respectively. As shown in Table 13, compared to the LBP based template matching in Section 5.1, AdaBoost (Boosted-LBP) provides improved performance. We also show the confusion matrix of 7-class recognition using AdaBoost in Table 14, where Disgust, Joy, Surprise and Neutral can be recognized with high accuracy. It can be seen that AdaBoost's performance is inferior to that of SVM (RBF) reported in Table 5 for most expressions except Fear and Neutral.

We further combine feature selection by AdaBoost with classification by SVM. In particular, we train SVM with the Boosted-LBP features. In each trial of the 10-fold cross-validation, we applied AdaBoost to learn the discriminative LBP histograms for each expression, and then utilized the union of the selected LBP histograms as the input for SVMs. For example, in Fig. 11, the union of all sub-regions selected resulted in a total of 51 LBP histograms. The generalization performance of Boosted-LBP based SVM is summarized in Table 15, where the degree of the polynomial kernel is 1 and the standard deviation for the RBF kernel is 2^{11}. For comparison, we also include the recognition performance of LBP based SVMs (in Section 5.2) in Table 15. We observe that Boosted-LBP based SVMs outperform LBP-based SVMs by around 2.5–3.5% points. The 7-class expression recognition result of 91.4% is very encouraging, compared to the state of the art [11]. Bartlett et al. [19] obtained the best performance 93.3% by selecting a subset of Gabor filters using AdaBoost and then training SVM on the outputs of the selected filters. With regard to the 6-class recognition,

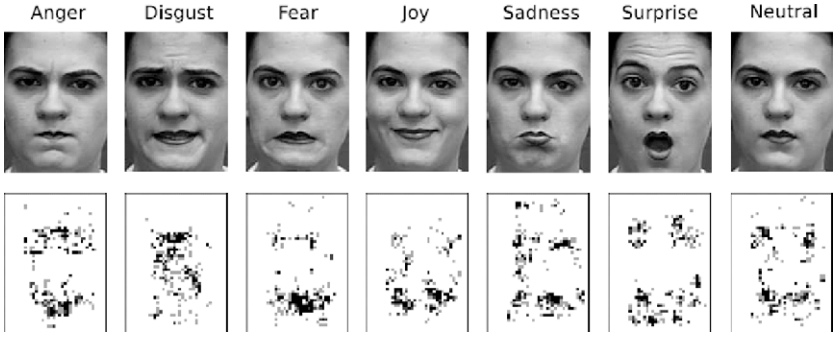

Fig. 10. Distributions of the top 50 sub-regions (LBP histograms) selected AdaBoost for each expression.

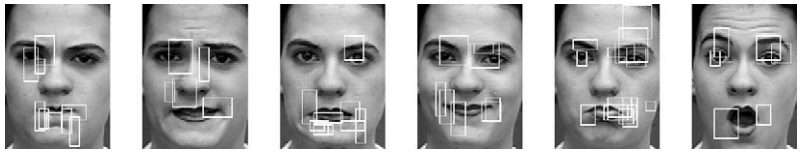

Fig. 11. The sub-regions (LBP histograms) selected by AdaBoost for each emotion. from left to right: Anger, Disgust, Fear, Joy, Sadness, Surprise.

Table 13
Recognition performance of Boosted-LBP vs LBP

	7-Class recognition (%)	6-Class recognition (%)
AdaBoost (Boosted-LBP)	85.0 ± 4.5	89.8 ± 4.7
LBP + template matching	79.1 ± 4.6	84.5 ± 5.2

Table 14
Confusion matrix of 7-class facial expression recognition using AdaBoost (Boosted-LBP)

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sadness (%)	Surprise (%)	Neutral (%)
Anger	**66.6**	3.7	2.0	0	7.3	0	20.4
Disgust	0	**92.5**	2.5	0	0	0	5.0
Fear	0	0	**70.0**	17.0	3.0	0	10.0
Joy	0	0	2.5	**90.1**	0	0	7.4
Sadness	6.4	0	0	0	**61.2**	0.8	31.6
Surprise	0	0	1.3	0	0.5	**92.5**	5.7
Neutral	0	0	0.8	0.4	3.6	0	**95.2**

Table 15
Recognition performance of Boosted-LBP based SVMs vs LBP based SVMs

	7-Class		6-Class	
	Boosted-LBP (%)	LBP (%)	Boosted-LBP (%)	LBP (%)
SVM (linear)	91.1 ± 4.0	88.1 ± 3.8	95.0 ± 3.2	91.5 ± 3.1
SVM (polynomial)	91.1 ± 4.0	88.1 ± 3.8	95.0 ± 3.2	91.5 ± 3.1
SVM (RBF)	91.4 ± 3.8	88.9 ± 3.5	95.1 ± 3.4	92.6 ± 2.9

the result of 95.1% is, to our best knowledge, the best recognition rate reported so far in the published literature on this database. Previously Tian [26] achieved 94% performance using a three-layer neural networks when combining geometric features and Gabor wavelet features. The confusion matrix of 7-class expression recognition using Boosted-LBP based SVM (RBF) is shown in Table 16. We can observe that, Disgust, Joy and Surprise can be recognized with very high accuracy (more than 97%), and Sad is the easiest confused expression with recognition accuracy around 75%. We also re-conducted the experiments on low-resolution face images in Section 6 using the Boosted-LBP features, and the recognition rates all increase 3–5%.

We also evaluated LDA using Boosted-LBP features. As discussed in Section 5.3, in each trial of the 10-fold cross-validation, the training data was first projected into a PCA subspace with 98% of variance kept, and the dimension of LDA subspace was $c-1$. The nearest-neighbor classifier was adopted as the classifier using the Euclidean distance measure. Boosted-LBP based LDA obtained the generalization performance of 77.6% for 7-class recognition and 84.2% for 6-class recognition. As shown in Table 17, LDA's performance is clearly improved by using Boosted-LBP features. But the performance of LDA is still inferior to that of SVM.

Table 16
Confusion matrix of 7-class facial expression recognition using Boosted-LBP based SVM

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sadness (%)	Surprise (%)	Neutral (%)
Anger	**85.1**	2.7	0	0	8.6	0	3.6
Disgust	0	**97.5**	0.8	1.7	0	0	0
Fear	0	1.0	**79.9**	11.0	3.1	1.0	4.0
Joy	0	0	0	**97.5**	0.4	0	2.1
Sadness	12.0	0	0.8	0	**74.7**	0	12.5
Surprise	0	0	1.3	0.9	0	**97.3**	0.5
Neutral	1.2	0	0.8	3.6	2.4	0	**92.0**

Table 17
Recognition performance of LBP based LDA vs Boosted-LBP based LDA

	7-Class recognition (%)	6-Class recognition (%)
LBP based LDA	73.4 ± 5.6	79.2 ± 7.2
Boosted-LBP based LDA	77.6 ± 5.7	84.2 ± 6.1

8. Generalization to other datasets

We evaluated the Boosted-LBP based SVM approach on another two publicly available databases: the MMI database [57] and the JAFFE database [7]. The MMI database includes more than 20 students and research staff members of both sexes (44% female), ranging in age from 19 to 62, having either a European, Asian, or South American ethnic background. Subjects were instructed to display 79 series of facial expressions, six of which are prototypic emotions. Image sequences have neutral faces at the beginning and the end, and were digitized into 720×576 pixels. Some sample images from the MMI database are shown in Fig. 12. Although the original data in the MMI database are color images, in our experiment, we converted them to 8-bit grayscale images. As can be seen, the subjects displayed facial expressions with and without glasses, which makes facial expression recognition more difficult. The JAFFE database consists of 213 images of Japanese female facial expression. Ten expressers posed 3 or 4 examples for each of the seven basic expressions (six emotional expressions plus neutral face). The image size is 256×256 pixels. Fig. 13 shows some sample images from the JAFFE database.

In our experiments, 96 image sequences were selected from the MMI database. The only selection criterion is that a sequence can be labeled as one of the six basic emotions. The sequences come from 20 subjects, with 1–6 emotions per subject. The neutral face and three peak frames of each sequence (hence, 384 images in total) were used for 7-class expression recognition. All 213 images of the JAFFE database were used for 7-class expression recognition. As we did on the Cohn–Kanade database, we normalized faces from the MMI database and the JAFFE database to a fixed distance be-

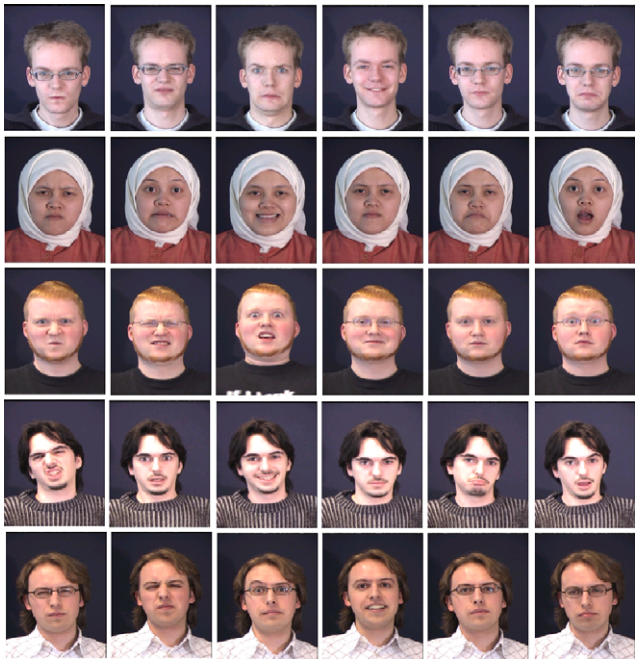

Fig. 12. The sample face expression images from the MMI database.

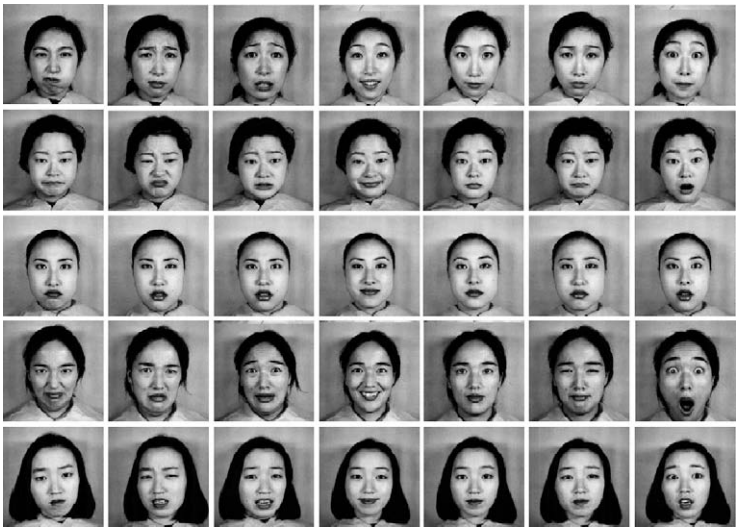

Fig. 13. The sample face expression images from the JAFFE database.

tween the two eyes; face images of 110 × 150 pixels were cropped from original frames based on the two eyes location.

We first performed 10-fold cross-validation on each dataset, and the recognition rates are shown in the top two rows of Table 18, where the degree of the polynomial kernel is 1 and the standard deviation for the RBF kernel is 2^{15} for the MMI database and 2^{8} for the JAFFE database. The best recognition performance of 86.9% on the MMI database is inferior to that on the Cohn–Kanade database. This is possibly because that there are fewer images in the dataset, and subjects are wearing glasses. The performance on the JAFFE is worst overall compared to that of the Cohn–Kanade database and the MMI database, and this may be also due to a much small dataset. With LBP features and the linear programming technique, Feng et al. [51] reported the performance of 93.8% on the JAFFE database. They preprocessed the images using the CSU Face Identification Evaluation System [58] to exclude nonface area with an elliptical mask. Liao et al. [34] recently reported the recognition performance of 85.6% on the JAFFE database, but they did not conducted 10-fold cross-validation.

We then performed across-dataset experiments, i.e., we performed LBP feature selecting and SVM training on the Cohn–Kanade database, and then tested the classifier on the MMI database and the JAFFE database, respectively. Recognition results are shown in the bottom two rows of Table 18, where the degree of the polynomial kernel is 1 and the standard deviation for the RBF kernel is 2^{14} for the MMI database and 2^{11} for the JAFFE database. We observe that generalization performance across datasets was much lower, such as around 50% on the MMI database and around 40% on the JAFFE database. These results actually reenforce Bartlett et al.'s recent finding [59], where they trained selected Gabor-wavelet features based SVMs on the Cohn–Kanade database and tested them on another Pictures of Facial Affect database, and obtained 56–60% performance. As we preprocessed face images of different databases in the same way, the only difference between them is that they were collected under different controlled environments. So the current expression classifier trained on a single dataset with uniformly controlled environment works well only within that dataset. In order to generalize across image collection environments, we have to collect large training datasets with variations in image conditions [59].

Table 18
Generalization performance of Boosted-LBP based SVM on other datasets

	SVM (linear) (%)	SVM (polynomial) (%)	SVM (RBF) (%)
MMI	86.7	86.7	86.9
JAFFE	79.8	79.8	81.0
Train:Cohn–Kanade Test:MMI	50.8	50.8	51.1
Train:Cohn–Kanade Test:JAFFE	40.4	40.4	41.3

9. Conclusions and future work

In this paper, we present a comprehensive empirical study of facial expression recognition based on Local Binary Patterns features. Different classification techniques are examined on several databases. The key issues of this work can be summarized as follows:

1. Deriving an effective facial representation from original face images is a vital step for successful facial expression recognition. We empirically evaluate LBP features to describe appearance changes of expression images. Extensive experiments illustrate that LBP features are effective and efficient for facial expression recognition.
2. One challenge for facial expression recognition is recognizing facial expressions at low resolutions, as only compressed low-resolution video input is available in real-world applications. We investigate LBP features on low-resolution images, and observe that LBP features perform stably and robustly over a useful range of low resolutions of face images.
3. We adopt AdaBoost to learn the most discriminative LBP features from a large LBP feature pool. Best recognition performance is obtained by using SVM with Boosted-LBP features. However, this method has limitation on generalization to other datasets.

Since the performance of the boosted strong classifier originates in the characteristics of its weak hypothesis space, we will evaluate other kinds of weak classifiers as alternative to template matching, in order to achieve better classification performance. One limitation of this work is that the recognition is performed by using static images without exploiting temporal behaviors of facial expressions. The psychological experiments by Bassili [60] have suggested that facial expressions are more accurately recognized from a dynamic image than from a single static image. We will explore temporal information in our future work. Recently volume

LBP and LBP from three orthogonal planes have been introduced for dynamic texture recognition [61], showing promising performance on facial expression recognition in video sequences. Another limitation of the current work is that we do not consider head pose variations and occlusions, which will be addressed in our future work. We will also study the effect of imprecise face location on expression recognition results.

Acknowledgements

We would like to thank Prof. Jeffery Cohn for the use of the Cohn–Kanade database, Prof. Maja Pantic and Dr. Michel F. Valstar for the use of the MMI database, and Dr. Michael J. Lyons for the use of the JAFFE database.

References

[1] M. Pantic, L. Rothkrantz, Automatic analysis of facial expressions: the state of art, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1424–1445.
[2] B. Fasel, J. Luettin, Automatic facial expression analysis: a survey, Pattern Recognition 36 (2003) 259–275.
[3] M. Pantic, L. Rothkrantz, Toward an affect-sensitive multimodal human–computer interaction, in: Proceeding of the IEEE, vol. 91, 2003, pp. 1370–1390.
[4] Y. Tian, T. Kanade, J. Cohn, Handbook of Face Recognition, Springer, 2005 (Chapter 11. Facial Expression Analysis).
[5] Y. Yacoob, L.S. Davis, Recognizing human facial expression from long image sequences using optical flow, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (6) (1996) 636–642.
[6] I. Essa, A. Pentland, Coding, analysis, interpretation, and recognition of facial expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 757–763.
[7] M.J. Lyons, J. Budynek, S. Akamatsu, Automatic classification of single facial images, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (12) (1999) 1357–1362.
[8] G. Donato, M. Bartlett, J. Hager, P. Ekman, T. Sejnowski, Classifying facial actions, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (10) (1999) 974–989.
[9] M. Pantic, L. Rothkrantz, Expert system for automatic analysis of facial expression, Image and Vision Computing 18 (11) (2000) 881–905.
[10] Y. Tian, T. Kanade, J. Cohn, Recognizing action units for facial expression analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2) (2001) 97–115.
[11] I. Cohen, N. Sebe, A. Garg, L. Chen, T.S. Huang, Facial expression recognition from video sequences: temporal and static modeling, Computer Vision and Image Understanding 91 (2003) 160–187.
[12] L. Yin, J. Loi, W. Xiong, Facial expression representation and recognition based on texture augmentation and topographic masking, in: ACM Multimedia, 2004.
[13] M. Yeasin, B. Bullot, R. Sharma, From facial expression to level of interests: a spatio-temporal approach, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2004.
[14] J. Hoey, J.J. Little, Value directed learning of gestures and facial displays, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2004.
[15] Y. Chang, C. Hu, M. Turk, Probabilistic expression analysis on manifolds, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2004.
[16] R.E. Kaliouby, P. Robinson, Real-time inference of complex mental states from facial expressions and head gestures, in: IEEE CVPR Workshop on Real-time Vision for Human–Computer Interaction, 2004.
[17] M. Pantic, L.J.M. Rothkrantz, Facial action recognition for facial expression analysis from static face images, IEEE Transactions on Systems, Man, and Cybernetics 34 (3) (2004) 1449–1461.
[18] Y. Zhang, Q. Ji, Active and dynamic information fusion for facial expression understanding from image sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (5) (2005) 1–16.
[19] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, J. Movellan, Recognizing facial expression: machine learning and application to spotaneous behavior, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005.
[20] F. Dornaika, F. Davoine, Simultaneous facial action tracking and expression recognition using a particle filter, in: IEEE International Conference on Computer Vision (ICCV), 2005.
[21] C.S. Lee, A. Elgammal, Facial expression analysis using nonlinear decomposable generative models, in: IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG), 2005.
[22] M. Valstar, I. Patras, M. Pantic, Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data, in: IEEE Conference on Computer Vision and Pattern Recognition Workshop, vol. 3, 2005, pp. 76–84.
[23] M. Valstar, M. Pantic, Fully automatic facial action unit detection and temporal analysis, in: IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2006, p. 149.
[24] M. Pantic, I. Patras, Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences, IEEE Transactions on Systems, Man, and Cybernetics 36 (2) (2006) 433–449.
[25] Z. Zhang, M.J. Lyons, M. Schuster, S. Akamatsu, Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron, in: IEEE International Conference on Automatic Face & Gesture Recognition (FG), 1998.
[26] Y. Tian, Evaluation of face resolution for expression analysis, in: CVPR Workshop on Face Processing in Video, 2004.
[27] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distribution, Pattern Recognition 29 (1) (1996) 51–59.
[28] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7) (2002) 971–987.
[29] T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, in: European Conference on Computer Vision (ECCV), 2004.
[30] A. Hadid, M. Pietikäinen, T. Ahonen, A discriminative feature space for detecting and recognizing faces, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2004.
[31] X. Feng, A. Hadid, M. Pietikäinen, A coarse-to-fine classification scheme for facial expression recognition, International Conference on Image Analysis and Recognition (ICIAR), Lecture Notes in Computer Science, vol. 3212, Springer, 2004, pp. 668–675.
[32] Y. Tian, L. Brown, A. Hampapur, S. Pankanti, A. Senior, R. Bolle, Real world real-time automatic recognition of facial expression, in: IEEE Workshop on Performance Evaluation of Tracking and Surveillance (PETS), Australia, 2003.
[33] C. Shan, S. Gong, P.W. McOwan, Robust facial expression recognition using local binary patterns, in: IEEE International Conference on Image Processing (ICIP), Genoa, vol. 2, 2005, pp. 370–373.
[34] S. Liao, W. Fan, C.S. Chung, D.-Y. Yeung, Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features, in: IEEE International Conference on Image Processing (ICIP), 2006, pp. 665–668.
[35] C. Darwin, The Expression of the Emotions in Man and Animals, John Murray, London, 1872.
[36] M. Suwa, N. Sugie, K. Fujimora, A preliminary note on pattern recognition of human emotional expression, in: International Joint Conference on Pattern Recognition, 1978, pp. 408–410.
[37] M. Bartlett, G. Littlewort, C. Lainscsek, I. Fasel, J. Movellan, Machine learning methods for fully automatic recognition of facial expressions and facial actions, in: IEEE International Conference on Systems, Man & Cybernetics, Netherlands, 2004.
[38] M. Turk, A.P. Pentland, Face recognition using eigenfaces, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1991.
[39] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
[40] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, Face recognition by independent component analysis, IEEE Transactions on Neural Networks 13 (6) (2002) 1450–1464.
[41] C. Shan, S. Gong, P.W. McOwan, Appearance manifold of facial expression, in: N. Sebe, M.S. Lew, T.S. Huang (Eds.), IEEE ICCV workshop on Human–Computer Interaction, Vol. 3723 of Lecture Notes in Computer Science, Springer, Beijing, 2005, pp. 221–230.
[42] C. Padgett, G. Cottrell, Representing face images for emotion classification, in: Advances in Neural Information Processing Systems (NIPS), 1997.
[43] P. Ekman, W. Friesen, Facial Action Coding System: A Technique for Measurement of Facial Movement, Consulting Psychologists Press, 1978.
[44] P. Ekman, W. Friesen, Pictures of Facial Affect, Consulting Psychologists, 1976.
[45] T. Kanade, J. Cohn, Y. Tian, Comprehensive database for facial expression analysis, in: IEEE International Conference on Automatic Face & Gesture Recognition (FG), 2000.
[46] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001.
[47] M.R. Everingham, A. Zisserman, Regression and classification approaches to eye localization in face images, in: IEEE International Conference on Automatic Face & Gesture Recognition (FG), 2006, pp. 441–446.
[48] M. Bartlett, G. Littlewort, I. Fasel, 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.
[49] V.N. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[50] C.-W. Hsu, C.-C. Chang, C.-J. Lin, A Practical Guide to Support Vector Classification, Tech. Rep., Taipei, 2003.
[51] X. Feng, M. Pietikäinen, T. Hadid, Facial expression recognition with local binary patterns and linear programming, Pattern Recognition and Image Analysis 15 (2) (2005) 546–548.
[52] G. Guo, C.R. Dyer, Simultaneous feature selection and classifier training via linear programming: a case study for face expression recognition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2003.
[53] R.E. Schapire, Y. Singer, Improved boosting algorithms using confidence-rated predictions, Maching Learning 37 (3) (1999) 297–336.

[54] G. Zhang, X. Huang, S.Z. Li, Y. Wang, X. Wu, Boosting local binary pattern (lbp)-based face recognition, in: Chinese Conference on Biometric Recognition (SINOBIOMETRICS), 2004, pp. 179–186.
[55] C. Shan, S. Gong, P.W. McOwan, Conditional mutual information based boosting for facial expression recognition, in: British Machine Vision Conference (BMVC), Oxford, vol. 1, 2005, pp. 399–408.
[56] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences 55 (1) (1997) 119–139.
[57] M. Pantic, M. Valstar, R. Rademaker, L. Maat, Web-based database for facial expression analysis, in: IEEE International Conference on Multimedia and Expo (ICME), 2005.
[58] D. Bolme, M. Teixeria, J. Beveridge, B. Draper, The CSU face identification evaluation system: its purpose, features and structure, in: International Conference on Vision Systems, 2003, pp. 304–311.
[59] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, J. Movellan, Dynamics of facial expression extracted automatically from video, Image and Vision Computing 24 (6) (2006) 615–625.
[60] J.N. Bassili, Emotion recognition: the role of facial movement and the relative importance of upper and lower area of the face, Journal of Personality and Social Psychology 37 (11) (1979) 2049–2058.
[61] G. Zhao, M. Pietikäinen, Dynamic texture recognition using local binary patterns with an application to facial expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (6) (2007) 915–928.