Facial Action Detection using Block-based Pyramid Appearance Descriptors

Bihan Jiang*, Michel F. Valstar† and Maja Pantic*‡
*Department of Computing, Imperial College London, UK
†Mixed Reality Lab, School of Computer Science, University of Nottingham, UK
‡Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Netherlands

***Abstract*—Facial expression is one of the most important non-verbal behavioural cues in social signals. Constructing an effective face representation from images is an essential step for successful facial behaviour analysis. Most existing face descriptors operate on the same scale, and do not leverage coarse v.s. fine methods such as image pyramids. In this work, we propose the sparse appearance descriptors Block-based Pyramid Local Binary Pattern (B-PLBP) and Block-based Pyramid Local Phase Quantisation (B-PLPQ). The effectiveness of our proposed descriptors is evaluated by a real-time facial action recognition system. The performance of B-PLBP and B-PLPQ is also compared with Block-based Local Binary Patterns (B-LBP) and Block-based Local Phase Quantisation (B-LPQ) . The system proposed here enables detection a much larger range of facial behaviour by detecting 22 facial muscle actions (Action Units, AUs), which can be practically applied for social behaviour analysis and synthesis. Results show that our proposed descriptor B-PLPQ outperforms all other tested methods for the problem of FACS Action Unit analysis and that systems which utilise a pyramid representation outperform those that use basic appearance descriptors.**

I. INTRODUCTION

Traditional computer interfaces usually emphasise the transmission of explicit messages whilst ignoring implicit information about the user. Thus the interaction could be forced and unnatural. The next generation of Human-Computer Interaction (HCI) and Interface will be able to perceive and understand human user's intentions and emotions as communicated by social and affective signals. Social signals are manifested through a multiplicity of non-verbal behavioural cues including facial expressions, body postures and gestures, vocal outbursts etc [12]. Therefore, automated analysis of non-verbal behaviour, and especially of expressive facial behaviour, is increasingly attracting attention.

The Facial Action Coding System (FACS) is the best known and most commonly used system developed for human observers to describe facial activities. The coding system defines atomic facial muscle actions called Action Units (AUs). With FACS, every possible facial expression (emotional or otherwise) can be described as a combination of AUs. For instance, an expression typically associated with happiness contains AU6 and AU 12 and sadness contains AU1, AU4 and AU15. As AUs are independent of interpretation, they can be used for any higher order decision making process such as cognitive states like interest and puzzlement, social behaviours like agreement and disagreement, social signals like status, trustworthiness and so on. Researchers have employed FACS to study everything from deception detection to data-driven for avatars. Recently, FACS has been adopted by computer animators in commercials.

A major factor which impedes the widespread use of FACS is the time required both to train human experts and to manually score the video tape [3]. An automatic Facial Action Recognition System (FARS) could vastly increase the amount of data a psychologist could analyse. As a tool for building an advanced user-interface, real-time performance is an essential property. Long delays make the interaction desynchronised and less efficient [12]. Hence it is crucial to find a trade-off between accuracy and efficiency. Though much progress has been made, robust real-time facial expression analysis remains difficult due to its subtlety, complexity, and variability [14].

Constructing an effective facial representation from face images is an essential step for successful facial expression recognition. Traditionally the feature extraction approaches may be divided into two streams: geometric feature-based methods and appearance-based methods. Geometric feature based methods employ the geometrical properties of a face. On the other hand, changes in image texture such as those created by wrinkles, bulges, and changes in feature shapes are captured by appearance based features. Typical examples include Gabor filters and Haar-like filters.

Local Binary Pattern (LBP) and Local Phase Quantisation (LPQ) are effective appearance descriptors which have been successfully applied in texture classification and face analysis. They are local appearance descriptors, which means they are able to capture subtle appearance changes. This is vital in facial expression recognition. Moreover, LBP is tolerant to illumination changes and efficient to compute, and LPQ is blur-invariant. These are desirable properties for real-time applications. The feasibility of LBP and LPQ for facial expression recognition has been shown in many existing works. Shan et al. [14] demonstrated promising performance in compressed low-resolution video sequences captured in real-world environments by using LBP. LBPs are also used to study multi-view facial expression recognition by Moore and Bowden [8]. Yang and Bhanu [17] won the Facial Recognition and Analysis challenge (FERA2011) - Emotion recognition sub-challenge - by using both LBP and LPQ features. LBP and LPQ have previously been applied to AU detection by the authors of this paper [6].

Pyramid transform is an effective multi-resolution analysis approach. There are a number of works extending the conven-

tional descriptors to the pyramid transform domain. Yang et al. [4] proposed a new pyramid Gabor features for emotion recognition. The work from [2] applied a pyramid Histogram of Oriented Gradients (PHOG) for smile recognition. Recently, Qian et al. [13] extended the conventional LBP to the pyramid transform domain called Pyramid Local Binary Pattern (PLBP) for texture classification and face recognition, which showed satisfactory performances with low computational costs.

However, as the PLBP descriptor is effectively a histogram of the PLBP patterns taken over the entire face image, it only captures global appearance statistics, thus removing all information regarding shape. To re-introduce a measure of shape in our appearance descriptor while maintaining its shift and scale robustness, we present an adapted version of PLBP called Block-based PLBP (B-PLBP), which we applied to the problem of AU detection. Similarly, B-PLPQ is proposed to investigate the merits of this representation.

The effectiveness of these descriptors is evaluated by a fully automatic AU detection system and tested on the expression data which are collected from the MMI database[15]. Fig. 1 shows an overview of the proposed system. In order to detect the upper and lower face AUs, we adopted Support Vector Machine (SVM) classifiers, one for each AU, which are trained on a subset of the most informative features selected by GentleBoost. To extract these appearance features, we first find the face in the input static image using an adapted version of the Viola and Jones face detector [16]. The C++ code of the face detector runs at about 500 Hz on a 3.2-GHz Pentium 4. Next the detected face images are registered to remove head rotations and scale variations by using the OpenCV implementation of an object detector to locate the eyes. Based on that, the face image is scaled to fix the distance between the eye locations, and then cropped to be 200 by 200 pixels. After that, the registered image is divided into small blocks and the B-LBP, B-LPQ, B-PLBP and B-PLPQ features are extracted. The histograms from all blocks are concatenated to form a single feature vector.

Our key contributions are threefold. First, we propose block-based pyramid local binary pattern and local phase quantisation (B-PLBP, B-PLPQ). Secondly, the proposed appearance descriptors are applied to AU detection. Finally, the experimental results show that B-PLPQ outperforms the three other descriptors for FACS AU analysis in terms of recognition accuracy.

The remainder of this paper is organised as follows. Section II briefly describes the basic principle of static appearance descriptors LBP, LPQ, and our proposed extensions B-PLBP and B-PLPQ. Section III presents the classification technique used in this work and the different kernels tested, while the evaluation procedures and test results are given in Section IV. Section V provides the conclusions of our research.

II. FEATURE EXTRACTION

Recognising facial expressions from static images is a more challenging problem than from image sequences, as less information about expressive actions is available. For example,

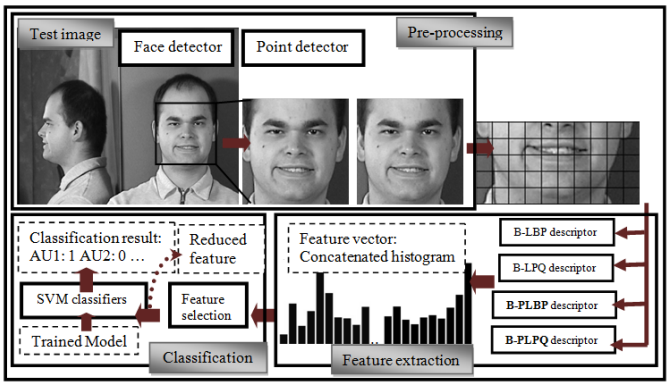

Fig. 1. The outline of our proposed system

without a neutral reference frame, it is impossible to tell from a still image whether the appearance of the eyebrows indicates a neutral expression, or that the brows are slightly raised. Still, often a single image can provide enough information for AU detection. In this section we explain how we combine the pyramid transform with a sub-division of the image space to derive our Block-based pyramid features.

A. Local Binary Patterns

Local Binary Patterns (LBP) were first introduced by Ojala et al. in [9], and proved to be a powerful means of texture description. The operator labels each pixel by thresholding a 3×3 neighbourhood of each pixel with respect to the centre value. Converting the 8-bit pattern to a decimal number, a 256-bin histogram of the LBP labels computed over a region is used as a texture descriptor.

A LBP 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 [10]. Using only uniform LBPs greatly reduces the length of the feature vector. The number of possible patterns for a neighbourhood of 8 pixels is 256 for the basic LBP but only 59 for uniform LBP. Hence, in this work we adopt uniform LBPs.

The occurrence of the uniform patterns over a region is recorded by a histogram. After applying the LBP operator to an image, a histogram of the labelled image $f(x, y)$ can be defined as

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

where n is the maximum label number produced by the LBP operator and $I(A)$ is the indicator function, which returns 1 if A is true, and 0 otherwise.

For B-LBP, a face is divided into $m \times n$ local regions, from which LBP histograms are extracted and then concatenated into a single, spatially enhanced feature histogram. Many extensions are made for conventional LBP descriptors. Refer to [5] for an extensive overview of LBP-based descriptors.

B. Local Phase Quantisation

The Local Phase Quantisation (LPQ) operator was originally proposed by Ojansivu and Heikkila as a texture descriptor that is robust to image blurring [11]. The descriptor uses local phase information extracted using the 2-D DFT or, more precisely, a short-term Fourier transform (STFT) computed over a rectangular M-by-M neighbourhood N_x at each pixel position $\mathbf{x}$ of the image $f(\mathbf{x})$ defined by

$$F(\mathbf{u}, \mathbf{x}) = \sum_{\mathbf{y} \in N_x} f(\mathbf{x}\text{-}\mathbf{y}) e^{-j2\pi \mathbf{u}^T \mathbf{y}} = \mathbf{w}_\mathbf{u}^T \mathbf{f}_\mathbf{x} \tag{2}$$

where $\mathbf{w_u}$ is the basis vector of the 2-D DFT at frequency $\mathbf{u}$, and $\mathbf{f_x}$ is the vector containing all M^2 samples from N_x.

The STFT is efficiently evaluated for all image positions $x \in \{x_1, ..., x_N\}$ using simply 1-D convolutions for the rows and columns successively. The local Fourier coefficients are computed at four frequency points: $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u_3 = [a, a]^T$, and $u_4 = [a, -a]^T$, where a is a sufficiently small scalar ($a = 1/M$ in our experiments). For each pixel position this results in a vector $F_x = [F(u_1, x), F(u_2, x), F(u_3, x), F(u_4, x)]$. The phase information in the Fourier coefficients is recorded by examining the signs of the real and imaginary parts of each component in F_x. This is done by using a simple scalar quantiser

$$q_j = \begin{cases} 1 & \text{if } g_j \geq 0 \text{ is true} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where $g_j(x)$ is the jth component of the vector $G_x = [\text{Re}\{F_x\}, \text{Im}\{F_x\}]$. The resulting eight bit binary coefficients $g_j(x)$ are represented as integers using binary coding:

$$f_{\text{LPQ}}(x) = \sum_{j=1}^{8} q_j 2^{j-1}. \tag{4}$$

As a result, a histogram of these values from all positions is composed and used as a 256-dimensional feature vector in classification.

It can be shown that in quantisation the information is maximally preserved if the samples to be quantised are statistically independent [11]. In practice, the neighbouring pixels are highly correlated in real images, which will lead to dependency between the Fourier coefficients g_j which are quantized in LPQ. Therefore Ojansivu et al. [11] improve LPQ by introducing a simple de-correlation mechanism. This is what we adopted in this work. For more details, please refer to [11]. The B-LPQ features are extracted in a similar way to LBP histograms from non-overlapping rectangular regions and concatenated into a feature vector.

C. Block-based pyramid representation

The block-based representation of local texture descriptors was first proposed by Ahonen et al. [1]. They divided face images into $m \times n$ local regions, from which LBP histograms can be extracted, and then concatenated them into a histogram. The resulting histogram encodes both the local texture and global shape of face images [5]. Also it is more robust to shift.

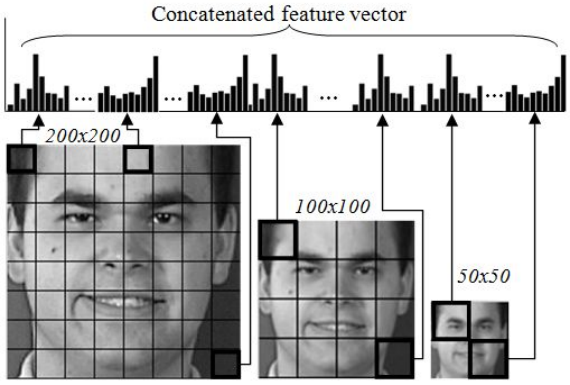

Fig. 2. The block-based pyramid representation

This scheme has been adopted in most existing studies for face representation (e.g. [6], [14]). However, some researchers are critical of this approach, suggesting that the subregions are not necessarily well aligned with facial features and the resulting facial description depends on the chosen window size and the position of these subregions [5].

Qian et al. [13] extended the conventional LBP to the pyramid transform domain. By cascading the LBP information obtains from a hierarchical spatial pyramid, PLBP takes texture resolution variations into account. They comprehensively compared PLBP with other LBP extensions for texture classification. They claimed that PLBP combines satisfactory performance with low computational cost. Different from texture analysis, the face can be seen as a composition of micro-patterns. However, a histogram computed over the whole image represents only the global distribution of the patterns thus the local information has been ignored.

Motivated by these ideas and in order to find a trade-off between robustness and sensitivity, we propose two novel descriptors B-PLBP and B-PLPQ which capture pixel-level, region-level and structure-level information for face representation. The face image is represented in an image pyramid by different spatial resolutions, e.g. 200×200, 100×100 and 50×50. Each pixel in the higher spatial pyramid levels is obtained by down sampling from its adjacent low-pass filtered high resolution image. Hence in the low resolution images, a pixel corresponds to a region in its high-resolution equivalent. For each pyramid level, the image is divided into regions. The dense appearance descriptor features extracted from each region, and each level of the pyramid, are concatenated into a single, spatially enhanced feature histogram (see Fig. 2). The final histogram is used as a feature vector to represent face image. A fixed size of window has been used. As shown in Fig. 2, the blocks in each level encodes different spatial information.

We have considered two ways of dividing regions in each pyramid level. One is to keep the number of blocks in each level fixed. However, that makes the block size smaller in each layer which, and results in capturing too detailed information. Therefore we decide to keep the size of each block constant.

In this configuration, three levels of locality are captured: the LBP/LPQ labels for histogram contains information about the patterns on a pixel level, the histogram computed over a block produces regional level and concatenating histograms from each region and each level encodes global and structural information.

In our experiments, a three level pyramid model and a region size of 25×25 pixels is used. That is, the different resolution face images are divided into 8×8, 4×4, 2×2 blocks respectively (see Fig. 2).

III. CLASSIFICATION

A previous successful technique to facial expression classification is Support Vector Machine (SVM). In this work, we adopted SVM as classifiers for AU detection. Given a training set of labelled examples $\{(x_i, y_i), i = 1, ..., l\}$, where $x_i \in R^n$ and $y_i \in \{1, -1\}$, a new test example x is classified by the following function:

$$f(x) = sgn(\sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x_i}, \mathbf{x}) + b) \quad (5)$$

where sgn function returns the sign of y, i.e. either 1 or -1, α_i are Lagrange multipliers of a dual optimisation problem that describe the separating hyperplane, $K()$ is a kernel function, and b is the threshold parameter of the hyperplane. Performing an implicit mapping of data into a higher dimensional feature space, which is defined by the kernel function, the training process is achieved by finding a linear separating hyper-plane with the maximal margin (M) to separate data in this higher dimensional space. The most frequently used kernel functions are the linear, polynomial, and Radial Basis Function (RBF). Recently, Maji et.al [7] proposed a histogram intersection kernel SVMs (IKSVMs). As our proposed descriptors are all histogram-based, we expect that this kernel is well-suited in our problem. To test this hypothesis, preliminary experiments are conducted. Results will be presented in Section IV-D.

The kernels we compared are the following:

- Linear kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$;
- Polynomial kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^d$;
- RBF kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2})$;
- Histogram intersection kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \sum_k min(\mathbf{X_i}(k), \mathbf{X_j}(k))$.

IV. EVALUATION

A. Training data selection

In [6], the authors proposed a heuristic approach to select data for training. It is noted that when more than one AU is activated, facial actions can appear very different from when they occur in isolation. For example, AU1 and AU2 pull the brow up, whereas AU4 pulls the brows together and down using primarily the corrugators muscle at the bridge of the nose. The appearance of AU4 changes dramatically depending on whether it occurs alone or in combination with AU1 and AU2. In order to capture the appearance of each action unit as fully as possible and thus build a richer data space, the heuristic approach takes in every video the first apex frames of each target AU, and all the apex frames where any other upper face AUs are in onset or offset (see Fig. 3). The shaded parts are the frames selected.

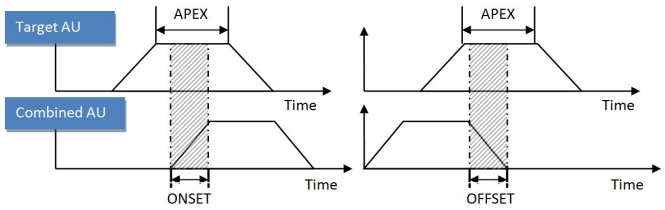

Fig. 3. The criterion of static data selection. The shaded areas are included in the dataset

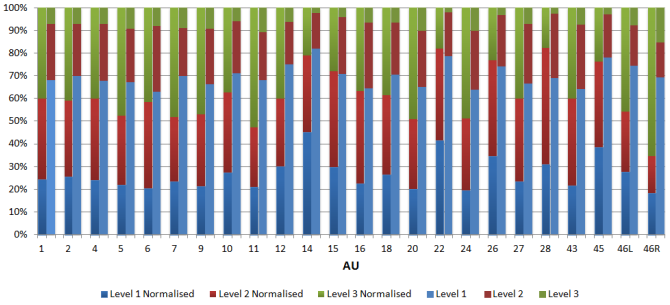

Fig. 4. The (normalised) percentage of feature selected from each pyramid level, trained on the entire MMI database (the level 1 has the highest resolution)

B. Feature selection

In order to select the most informative features, the GentleBoost algorithm is employed as a feature selector preceding each classifier. At each stage a weak classifier is trained on a subset of the data consisting of a single feature and iteratively boosted to a strong classifier of higher accuracy. At each iteration, the weak classifier which minimises the weighted error rate is selected, and the feature that this weak classifier represents is added to the list of selected features.

Fig. 4 illustrates the distribution of feature selected from each pyramid level. As we expected, most features are selected from the level with most detail and fewer features from the low resolution images. This can be explained in two ways: the higher resolution images result in more candidate features (as shown in the normalised distribution) , and they potentially capture more detail. Still, there are features selected even from the highest pyramid level for all tested AUs.

C. Comparison Setup

We evaluated the four appearance descriptors on 442 videos taken from the MMI database. In order to compare different approaches, the same evaluation process is performed. As this is a user independent system for FACS AU detection, the evaluation is done in a subject independent manner. Generalisation to new subjects is tested using 10-fold cross validation, where all videos are divided into ten subsets without mixing subjects.

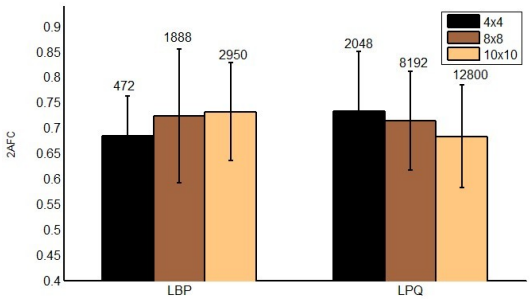

Fig. 5. The mean and standard deviation of the 2AFC scores over upper face AUs by using different block sizes tested on the MMI database. The number above each bar indicates the dimensionality of the features

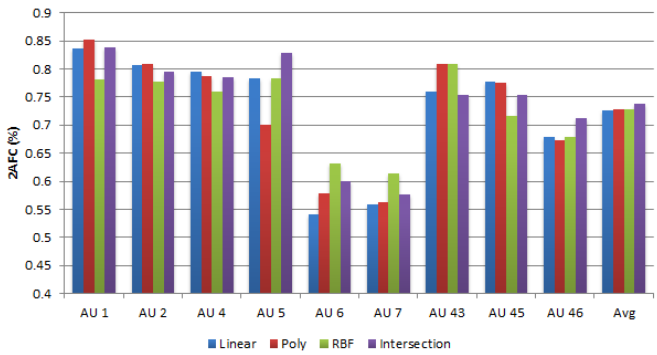

Fig. 6. The 2AFC (%) over all videos based on different SVM kernels, tested on the MMI database

At each iteration, nine of them are used to create training set and one is testing set used for testing. Hence no data from a subject appears in both the training and testing set.

The performance measure used in this work is the area under the ROC curve. By using the signed distance of each sample to the SVM hyper-plan and varying a decision threshold, we plot the hit rate (true positives) against the false alarm rate (false positives). The area under this curve is equivalent to percent correct in a 2-alternative forced task (2AFC) [3], which can be computed efficiently.

For each AU, we use only features extracted from either the upper or lower face, depending on whether the target AU causes appearance changes in the upper or lower face.

D. Experimental results

1) Block size optimisation: We conducted preliminary experiments to optimise the block size for B-LBP and B-LPQ. As shown in Fig 5, grid sizes of 4×4, 8×8 and 10×10 were tested. Unsurprisingly, for B-LBP, better results are attained when employing smaller block sizes. This is because subtle changes can be captured. This trend contrasts sharply with that of LPQ, where fewer grids produce better results. The reason for that lies in the way the LPQ features are extracted. As opposed to LBP, the local phase information is extracted after applying a Fourier Transform which means that even in a large region, the local information is still preserved.

2) Kernel functions: Fig 6 shows the average 2AFC scores performed with LBP based on different SVM kernels as discussed in Section III. The same selected B-LBP features were used for each kernel, and we used the heuristic approach training instance selection method. For all kernels, the parameters are optimised using cross-validation on the training set. Overall, the best results were reached with the histogram intersection kernel. This is to be expected as all the features used in this work are histogram-based. It is also worth pointing out the computational simplicity of IKSVM. In [7], the authors have shown IKSVM gives comparable accuracy while being $50\times$ faster and requiring $200\times$ less memory than the standard SVM implementation in their experiments.

3) Appearance descriptors: Fig. 7 presents the 10-fold cross-validation results using LBP, LPQ, PLBP, B-PLBP and B-PLPQ for 23 upper and lower face AUs. These 23 AUs are targeted because they have more than 10 sequences activated in the dataset. Note that in the manual labelling of the dataset, AU46 (wink) has been split up into 46L and 46R as the appearance differs greatly depending on whether using left or right eye. Hence it is treated as two separate AUs in our experiments.

To report the best performance of all systems, the histogram intersection kernel SVM is adopted in these experiments. We can see that, overall speaking, B-PLPQ produces best results among these four descriptors and the systems that utilise pyramid representation outperform those using basic appearance descriptors. The merit of pyramid representation is more obvious in the lower face AUs. One possible explanation is that the mouth movement usually causes a larger area of appearance changes in the lower face. For example, AU15 is lip corner depressor. It does not only change the shape of lips, but also produce appearance changes below the lip corners and on the chin boss. In this case, the spatial relation is essential. This is also true for AU14 (dimpler), AU20 (lip stretcher), and AU26 (jaw drop), which benefit most from the pyramid representation. On opposite to that, for instance, AU18 (lip pucker) and AU43 (eye closure), the appearance changes are more centralised around the mouth and eye area respectively. Thus local features are enough to capture all the changes.

4) Computational complexity: We also consider the computational complexity of the tested descriptors. All the algorithms are run in MATLAB environment on a PC (intel(R), core(TM)i7, CPU Q870 with 2.93 GHz, 8GB RAM). The average computational cost of B-LBP, B-LPQ, B-PLBP and B-PLPQ is 0.135s, 0.1526s, 0.412s and 0.453s respectively in computing an image of 200×200. The grid size is 25×25 for all descriptors and the pyramid level is 3.

This is not a fair comparison as it largely depends on the implementation. So Instead of comparing their computational time directly, we also analyse their complexity level. For an image of size $N \times N$, the complexity for LBP is $O(N^2)$. As we know, LPQ employs 2D STFT. The computational

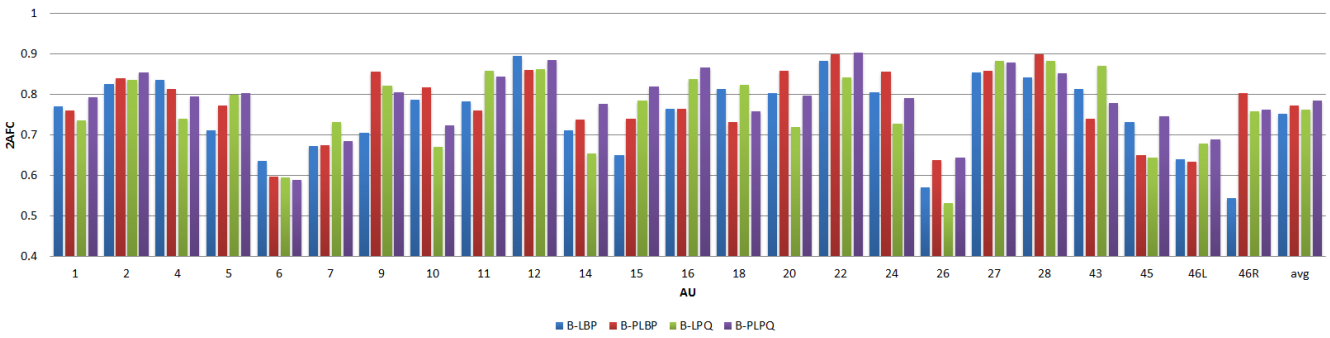

Fig. 7. The 2AFC (%) over all videos from the MMI database by using B-LBP, B-LPQ, B-PLBP and B-PLPQ

complexity of 1D FFT is $O(N^2 lgN)$. Hence to compute LPQ features from a $N \times N$ image, the computational complexity is $O(N^4 lgN)$. Hence LBP and LPQ are both in the same complexity class P. For their pyramid representation, by summing the computational cost from each pyramid level, assuming the number of levels is infinite, let C be the complexity of LBP/LPQ, then the complexity of a block-based pyramid descriptor is $O(C/1-q)$, where q is the common ratio of number of blocks between two successive levels $(0 < q < 1)$ and $1-q$ is a constant. So the pyramid versions of LBP/LPQ do not change the complexity level of the original descriptor.

V. CONCLUSIONS

We successfully implemented a robust and real-time AU detection system, based on the appearance descriptors B-LBP, B-LPQ and their pyramid extensions B-PLBP and B-PLPQ. Results show that systems based on LPQ generally achieve higher accuracy rate than those based on LBPs, and that systems that utilise a pyramid representation outperform those that don't. Although the family of block-based pyramid descriptors are more computationally expensive than the basic ones, they attain a higher recognition performance. All in all, the experimental results clearly show that our proposed descriptor B-PLPQ outperforms all other tested methods for the problem of FACS Action Unit analysis. The fastest of the systems described in this work, i.e. the LBP-based AU detector, is freely available as part of the SEMAINE framework, which can be downloaded from http://semaine.opendfki.de/.

ACKNOWLEDGMENTS

This work has been funded in part by the European Community's 7th Framework Programme [FP7/20072013] under the grant agreement no 231287 (SSPNet). The work of Michel Valstar was also funded in part by the EPSRC grant EP/H016988/1: Pain rehabilitation: E/Motion-based automated coaching. The work of Maja Pantic is further funded in part by the European Research Council under the ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB).

REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In *European Conference on Computer Vision*, pages 469–481, 2004.

[2] Y. Bai, L. Guo, L. Jin, and Q. Huang. A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition. In *Proceedings of the 16th IEEE international conference on Image processing*, pages 3269–3272, 2009.

[3] M. Bartlett, G. Littlewort-Ford, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behaviour. In *IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, pages 223–230, 2006.

[4] D. duan Yang, L. wen Jin, J. xun Yin, L. xin Zhen, and J. cheng Huang. Facial expression recognition with pyramid gabor features and complete kernel fisher linear discriminant analysis. *International Journal of Information Technology*, 11(9):91–100, 2005.

[5] D. Huang, C. Shan, and M. Ardabilian. Local binary pattern and its application to facial image analysis: A survey. *IEEE Trans. Systems, Man and Cybernetics, Part C*, 41:1–17, 2011.

[6] B. Jiang, M. F. Valstar, and M. Pantic. Action unit detection using sparse appearance descriptors in space-time video volumes. In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition (FG'11)*, 2011.

[7] S. Maji, A. Berg, and J. Malik. Classification using intersection kernel support vector machines is efficient. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[8] S. Moore and R. Bowden. Local binary patterns for multi-view facial expression recognition. *Computer Vision and Image Understanding*, 115(4):541–558, 2011.

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

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

[11] V. Ojansivu and J. Heikkila. Blur insensitive texture classification using local phase quantization. In *In Proc. Int. Conf. on Image and Signal Processing*, volume 5099, pages 236–243, 2008.

[12] M. Pantic, R. Cowie, F. D'ericco, D. Heylen, M. Mehu, C. Pelachaud, I. Poggi, M. Schroder, and A. Vinciarelli. *Social Signal Processing: The Research Agenda*, pages 511–538. Springer, 2011.

[13] X. Qian, X. Hua, P. Chen, and L.Ke. Plbp: An effective local binary patterns texture descriptor with pyramid representation. *Semi-Supervised Learning for Visual Content Analysis and Understanding*, 44(10):2502–2515, 2011.

[14] C. Shan, S. Gong, and P. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816, 2008.

[15] M. Valstar and M. Pantic. Induced disgust, happiness and surprise: an addition to the mmi facial expression database. In *Proc. Int'l Conf. Language Resources and Evaluation, W'shop on EMOTION*, pages 65–70, 2010.

[16] P. Viola and M. Jones. Robust real-time object detection. In *International Journal of Computer Vision*, 2010.

[17] S. Yang and B. Bhanu. Facial expression recognition using emotion avatar image. *Image Rochester NY*, page 866871, 2011.