



# Local binary patterns for multi-view facial expression recognition

S. Moore<sup>\*</sup>, R. Bowden

Centre for Vision Speech and Signal Processing University of Surrey, Guildford GU2 7JW, UK

## ARTICLE INFO

### Article history:

Received 22 August 2010

Accepted 8 December 2010

Available online 7 January 2011

### Keywords:

Facial expression recognition

Multi-view facial expression recognition

Head pose estimation

Local binary patterns

Local gabor binary patterns

## ABSTRACT

Research into facial expression recognition has predominantly been applied to face images at frontal view only. Some attempts have been made to produce pose invariant facial expression classifiers. However, most of these attempts have only considered yaw variations of up to 45°, where all of the face is visible. Little work has been carried out to investigate the intrinsic potential of different poses for facial expression recognition. This is largely due to the databases available, which typically capture frontal view face images only. Recent databases, BU3DFE and multi-pie, allows empirical investigation of facial expression recognition for different viewing angles. A sequential 2 stage approach is taken for pose classification and view dependent facial expression classification to investigate the effects of yaw variations from frontal to profile views. Local binary patterns (LBPs) and variations of LBPs as texture descriptors are investigated. Such features allow investigation of the influence of orientation and multi-resolution analysis for multi-view facial expression recognition. The influence of pose on different facial expressions is investigated. Others factors are investigated including resolution and construction of global and local feature vectors. An appearance based approach is adopted by dividing images into sub-blocks coarsely aligned over the face. Feature vectors contain concatenated feature histograms built from each sub-block. Multi-class support vector machines are adopted to learn pose and pose dependent facial expression classifiers.

© 2010 Elsevier Inc. All rights reserved.

## 1. Introduction

Over the last 20 years there has been a growing interest in improving the interaction between human and computers. As computing becomes more ubiquitous, human computer interaction will become more important. It can be argued that for a truly effective human computer interface, computers should be able to interact naturally with the user, in the same way that humans interact with other humans. In human to human interaction, Mehrabian [15] discovered that verbal cues provide 7% of the meaning of the message; vocal cues, 38%; and facial expressions, 55%. Thus facial expression provides more information about the interaction than the spoken words.

Ekman and Friesen [3] carried out research that indicates facial expressions are universal and innate. Ekman observed that members of an isolated tribe in Papua New Guinea could reliably identify the expressions of emotions in photographs. These expressions are commonly called the basic expressions which include *joy*, *surprise*, *anger*, *fear*, *sadness* and *disgust*.

Extracting an efficient representation of the face from images is an important step for successful facial expression recognition. In general, there are two common types of features used for facial

expression recognition: geometric based methods and appearance based methods [30]. Geometric features contain information about the location and shape of facial features. Appearance based features examine the appearance change of the face (including wrinkles, bulges and furrows) and are extracted by image filters applied to the face or sub regions of the face. Geometric features are sensitive to noise and tracking errors. Appearance based features are less reliant on initialization and can encode micro patterns in skin texture that are important for facial expression recognition. But appearance features do not generalize as well across individuals as they encode specific appearance information. In this paper we investigate appearance based features using large varied datasets and machine vision learning approaches to generalize across individuals and pose.

Psychophysical studies in saccadic eye movements [17] indicate that local appearance is important for classification. People can recognize objects when they seek regions where discriminating information is located. Our approach utilizes this finding by dividing face images into sub-blocks and comparing the similarities between these sub-blocks. This is a proven method for accurate facial expression recognition [4,24].

LBPs have been demonstrated to be successful as a texture descriptor in many computer vision applications [24,2,28]. One of the most important properties of LBPs are its tolerance to illumination change. Also the computational simplicity of the operator is a significant advantage over other approaches. In this paper we

<sup>\*</sup> Corresponding author.

E-mail addresses: [stephen.moore@surrey.ac.uk](mailto:stephen.moore@surrey.ac.uk) (S. Moore), [r.bowden@surrey.ac.uk](mailto:r.bowden@surrey.ac.uk) (R. Bowden).

investigate many variants of LBPs for multi-view facial expression recognition. Using different variants of a LBP allows us to investigate the importance of multi-resolution and orientation analysis for feature representation.

Facial expression databases primarily capture frontal view face images only [8,13]. Recent databases such as the BU3DFE and multi-pie databases allow an investigation of how a change in yaw viewing angle can effect facial expressions recognition. The BU3DFE database consists of 3D textured face models and motivates us to investigate the effects of pose change on facial expression recognition. We attempt to classify each of the prototypical expressions at five different yaw angles (0°, 30°, 45°, 60° and 90°). Another recent database multi-pie allows us to validate these experiments on real image data for seven different poses (0°, 15°, 30°, 45°, 60°, 75° and 90° yaw angles).

The remainder of this paper is as follows. In the next section background work is presented. In Section 3 a discussion and comparison of popular facial expression databases is presented. Section 4 introduces (LBPs) and some extensions including multi-scale local binary patterns ( $LBP^{ms}$ ) and local gabor binary patterns (LGBP). Section 5 presents experiments on the BU3DFE Dataset and a discussion on the findings. An investigation is presented into the effects of pose on overall facial expression recognition and individual expressions. Section 6 attempts to validate similar experiments on a live dataset using the recently released multi-pie dataset. We also investigate a local versus global approach for building feature histograms of the face. Frontal and profile face detectors are used to extract the face region. This allows us to investigate the tolerance of features to misalignment errors introduced by the face detectors. In Section 7 conclusions are drawn.

## 2. Background

Frequently used databases for facial expression recognition typically capture data at near frontal view [8,13]. High recognition rates for prototypical facial expressions have been recorded for these databases, in constrained settings [9,4,24]. Pose is one constraint that has largely been unexplored. This is mainly due to a lack of suitable data. Research in psychology has shown how pose can effect a humans ability to perceive facial expressions. Experiments using a Japanese noh mask, show that slight variations in pitch angle changes the two dimensional location of salient facial features which viewers misinterpret as non ridged changes due to muscle action [14]. Psychology experiments have shown that even a 15° yaw head pose change, results in statistically significant changes in how humans perceive emotion [18].

It is natural to assume that frontal pose is optimal for facial expression recognition, as at this view the whole face is visible. However, experiments in psychology suggest the optimal view for face recognition is 45°. While, computer vision experiments have shown similar findings for face recognition [12], little work has been carried out to investigate optimal view for facial expression recognition. Some work has investigated facial expression recognition for large head pose changes [25] and Pantic and Patras [22] explored recognition of facial action units from profile face image sequences. Wang et al. [27] show how sensitive 2D facial expression recognition approaches are to head pose variations. But this highlights the need to further investigate how head pose effects facial expression classifiers. Other approaches to pose invariant facial expression recognition, learn models of the whole face. These approaches typically do not consider views greater than 45°, when part of the face is occluded [23,1].

A recent database BU-3DFE [29] has initiated research into multi-view facial expression recognition. Hu et al. [6] focuses on facial expression recognition using LBPs, Histograms of Oriented

Gradients (HOGs) and the Scale Invariant Feature Transform (SIFT) to characterize facial expressions over five yaw rotation angles from frontal to profile views. Other contributions of this work are the strong performance increase when features are combined with Locality Preserving Projection (LPP). In [7], Hu et al. utilize the geometric 2D displacement of manually labeled facial points, and concatenate them together to form a feature vector as input to a SVM classifier. The main conclusion of [7] is that non-frontal views are better than frontal view for a computer to recognize facial expressions. An interesting question is if this conclusion is related to the geometric features used. In this paper, we explore this question using an appearance based approach. Limitations in the work of [6,7] are that features are extracted using a set of sparse manually labeled feature points. However it is not obvious how this approach can be applied to live captured data, since some feature points are not visible for large pose variations. We adopt a dense uniform sampling and use a multi-class support vector machine to learn pose and pose dependent facial expression classifiers.

## 3. Comparison of databases

Popular databases for facial expression recognition include the Cohn-kanade database [8] and the JAFFE database [13]. The JAFFE database contains 213 images of seven facial expressions (six basic facial expressions + one neutral) posed by 10 Japanese female models. Disadvantages of this database are the small number of subjects and the limited variability of the dataset. Subjects are all the same gender and the same ethnicity, thus reducing the complexity of the problem. Results of up to 95% have been reported for this database [9]. The Cohn-kanade database is the most popular dataset used for facial expression recognition. This database has sufficient subjects, different genders and ethnicities. However not all subjects perform all the basic expressions thus person independent experiments are not feasible as this would reduce significantly the number of subjects for experiments. Thus most experiments on this dataset are not person independent. Results of up to 93% have been achieved using gabor features [11] and  $LBP^{ms}$  [24].

The above mentioned databases do not allow investigation of how different view points effect facial expression recognition. The BU3DFE database provides 3D textured models of different facial expressions, from which images can be extracted by projecting 2D images at different yaw angles. An interesting characteristic of this database is that facial expressions are captured at different intensities. This allows an investigation into how different intensities effect recognition accuracy. Experiments in Section 5 use images for five different poses from the BU3DFE dataset to evaluate the effect of pose on facial expression recognition.

Another issue with the above databases is the lack of real world variabilities such as facial hair (beards and mustaches) and glasses. Fig. 1 shows examples of these variations in the recent multi-pie dataset. The top row of Fig. 1 shows examples of facial hair present in the multi-pie dataset. According to the Vision Council of America, About 64% of americans wear eyeglasses. Yet little research has been carried out for facial expression analysis with significant data for subjects wearing glasses, particularly for multi view facial expression recognition. In experiments 100 subjects are used where 49 subjects wore glasses in some or all of the sessions. An open question is how glasses effect multi-view facial expression recognition. The second column of Fig. 1 highlights some problems introduced by glasses. Picture D shows how glasses can occlude the shape of the eyebrows. Specular reflection can occur when subject wear glasses (picture E). Also at some angles glasses can occlude important facial features as seen in picture F of Fig. 1. These problems provide a challenging dataset which will test the robustness of the approach presented in this paper to such variations. Using

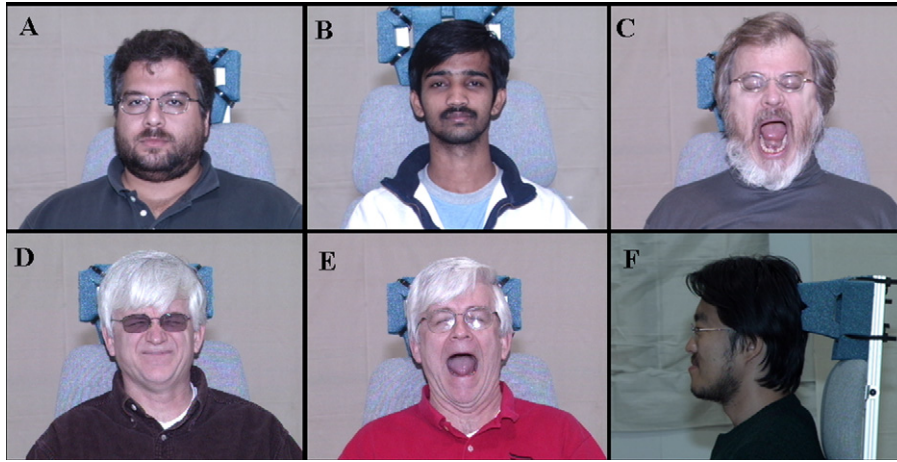


Fig. 1. Example of variations present in the multi-pie database. Top row – subjects with facial hair. Bottom row – occlusion due to subjects wearing glasses.

a similar approaches to the state of the art approaches ([11,24]) for a more difficult database will give a better reflection of performance.

Experiments presented in Section 5 use images from the BU3DFE database. Images are re-projected from a 3D textured model for different yaw angles and thus there is less variability in the synthesized dataset than in a live captured dataset. Also the 3D models used in the BU3DFE dataset were cropped models. The multi-pie dataset is used in Section 6 to validate the performance of the approach on a real dataset that addresses subject variability due to glasses and multiple recording sessions.

#### 4. Local binary patterns

The LBP operator was first introduced by Ojala et al. [19]. The operator labels the pixels  $f_p$  ( $P = 0, \dots, 7$ ) of an image by thresholding a  $3 \times 3$  neighborhood of each pixel with the value of the center pixel  $f_c$  and considering the result as a binary number  $S(f_p - f_c)$ . An example of the LBP operator is shown in Fig. 2.

$$S(f_p - f_c) = \begin{cases} 1 & \text{if } f_p \geq f_c \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Then, by assigning a binomial factor  $2^P$  for each  $S(f_p - f_c)$  the LBP is computed as follows:

$$LBP = \sum_{p=0}^7 S(f_p - f_c) 2^p \quad (2)$$

LBP's have proven to be very effective for image representation having been applied to visual inspection, motion detection and outdoor scene analysis. The most important properties of LBP features are their tolerance against monotonic illumination changes and their computational simplicity. The LBP operator detects many different

texture primitives (spot, line end, edge, corner), typically accumulated into a histogram over a region to capture local texture information.

Ojala et al. [20] extended this operator to use neighborhoods of different sizes, to capture dominant features at different scales. Notation  $LBP(P, R)$  denotes a neighborhood of  $P$  equally spaced sampling points on a circle of radius  $R$ . Fig. 2 shows a basic LBP where  $P = 8$  and  $R = 1$ . Ojala et al. [20] also showed that a small subset of the  $2^P$  patterns accounted for the majority of the texture of images, over 90% of all patterns for  $LBP(8, 1)$ . These patterns, called uniform patterns, contain at most two bitwise transitions from 0 to 1 or vice versa for a circular binary string. For example 01100000 and 11011111 are uniform patterns. These binary patterns, can be used to represent texture primitives such as spot, flat area, edge and corner. The uniform patterns contain in total  $(P - 1)P + 2$  binary patterns. Where  $(P - 1)P$  are rotational patterns, including edges and two non-rotational patterns, spot and flat area. Patterns where  $U(x) > 2$ , are defined as non-uniform patterns:

$$LBP_{P,R}^{u2} = \begin{cases} z & \text{if } U(LBP_{P,R}) \leq 2, LBP_{P,R} = I_z, I_z \in I, \\ & \text{where } |I| = (P - 1)P + 1 \\ (P - 1)P + 2 & \text{otherwise} \end{cases} \quad (3)$$

where,

$$U(LBP_{P,R}) = ||S(f_{p-1} - f_c) - S(f_0 - f_c)|| + \sum_{p=0}^P ||S(f_p - f_c) - S(f_{p-1} - f_c)|| \quad (4)$$

Uniform patterns with a  $U$  value of at most 2 are defined by superscript  $u2$  shown in Eq. (3). If  $U(x)$  is less than 2, the pixel is labeled using an index function  $I(z)$ . Otherwise it is labeled as non-uniform

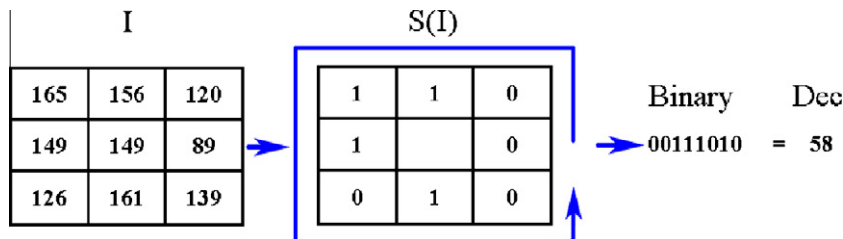


Fig. 2. The basic LBP operator. First the values in each pixel around the center pixel are thresholded with the center pixel. A binary number is extracted and a decimal value is calculated.



and assigned a value of  $(P - 1)P + 2$ .  $I(z)$  the indexing function, contains  $(P - 1)P + 2$  indices to assign to each uniform pattern. Using uniform patterns for a neighborhood where  $P = 8$ , reduces the histogram from 256 to 59 bins (58 bins for uniform patterns and 1 bin for non-uniform patterns). Fig. 3 shows the  $LBP^{u2}$  maps for each of the basic facial expressions from the multi-pie database. Experiments from [24] show that the performance of standard LBPs are similar to that of  $LBP^{u2}$  for frontal facial expression recognition.

#### 4.1. Rotation invariant LBP

Other extensions of the LBP operator used in this paper are rotation invariant LBP ( $LBP^r$ ) and rotation invariant uniform LBP ( $LBP^{riu2}$ ) [20]. To remove the effect of rotation i.e. to assign a unique identifier to each rotationally invariant LBP:

$$LBP_{P,R}^r = \min_{i=0}^{P-1} \{ROR(LBP_{P,R}, i)\} \quad (5)$$

Where  $ROR(x, i)$  performs a circular bitwise right shift on the  $P$ -bit number  $x$ ,  $i$  times. This operation further reduces the histogram, e.g.  $P = 8$   $LBP^r$  has 36 unique rotational invariant patterns. The performance of  $LBP^r$  features varies. Some patterns sustain rotation well, while other patterns do not and thus confuse the analysis [20]. The concept of uniform patterns can be extended to this feature, also reducing the number of bins from 36 to 9. This provides uniform rotational invariant local binary patterns  $LBP^{riu2}$ .

#### 4.2. Magnitude LBP

To further characterize the image information, the LBP operator is applied to the gradient magnitude image to produce a ( $LBP^{gm}$ ) image. To produce the gradient magnitude image, the image gradient for both  $x$  and  $y$  directions must be calculated. This can be achieved by using a first order derivative like the sobel operator. The sobel operator calculates the gradient of the image intensity at each point. The resulting map shows how smooth or sudden the image intensity changes at that point. The operator uses two  $3 \times 3$  kernels which are convolved with the original image to calculate approximations of the derivatives:

$$I_x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * I \quad \text{and} \quad I_y = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * I \quad (6)$$

Where  $*$  is the convolution operation.  $I$  is the source image,  $I_x$  and  $I_y$  are the two images where each point contains the horizontal and vertical derivative approximations. For each point the gradient approximations can be formulated to give the gradient magnitude image:

$$I_{gm} = \sqrt{I_x^2 + I_y^2} \quad (7)$$

$LBP^{u2}$  is applied to  $I_{gm}$  to create the  $LBP^{gm}$  feature map. This approach is a derivative based LBP which encodes the magnitude of local variation. Similar features have been successfully applied to frontal facial expressions recognition [9]. Fig. 4 shows the steps taken to produce a  $LBP^{gm}$  map.

Over a region, LBPs are accumulated in a histogram and the concatenation of these neighborhoods are then used as a descriptor. This characterizes the spatial structure of the local image texture.

All features mentioned above can be concatenated into a single feature vector HG for image  $LBP^{xxx}$ , with  $n$  sub-blocks:

$$HG(LBP^{xxx}) = (H_0, H_1, \dots, H_{n-1}). \quad (8)$$

where the histogram of the  $r$ th sub-block of  $LBP^{xxx}$  is computed by:

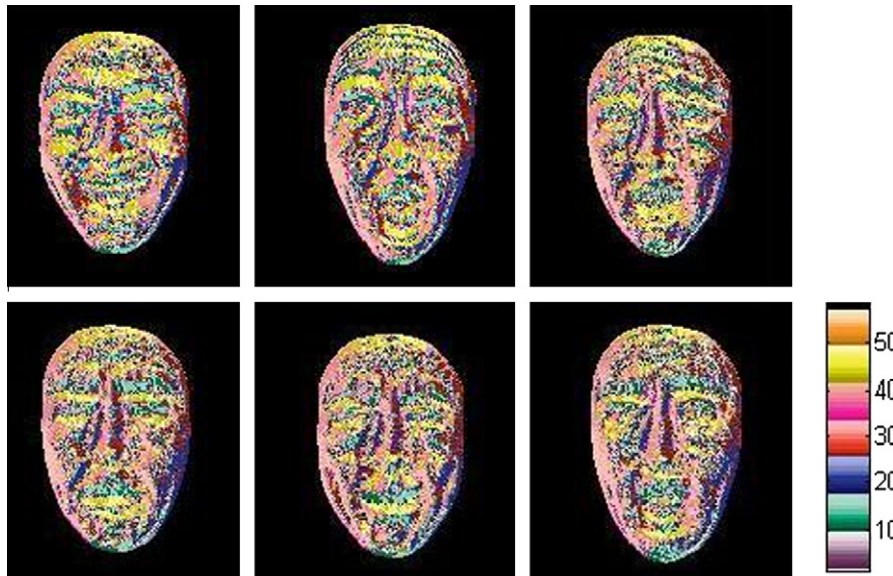
$$H_r = (h_{r,0}, h_{r,1}, \dots, h_{r,u-1}) \quad (9)$$

where  $u$  is the total number of bins for feature  $LBP^{xxx}$  and  $h$  is defined as:

$$h_i = \sum_{x,y} I\{LBP^{xxx}(x,y) = i\}, \quad i = 0, 1, \dots, u - 1 \quad (10)$$

where  $i$  is the  $i$ th bin of histogram  $h$ ,  $h_i$  is the number of patterns in the image with  $LBP^{xxx}$  pattern  $i$  and

$$I(A) = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{otherwise} \end{cases} \quad (11)$$



**Fig. 3.** Visualization of  $LBP^{u2}$  images for the six different facial expressions. Top row left to right, Joy, Surprise and Fear. Bottom row left to right, Sadness, Disgust and Anger. (white – dark spot, black – bright spot, gray – non-uniform patterns and other colors represent uniform patterns for different rotational angles). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

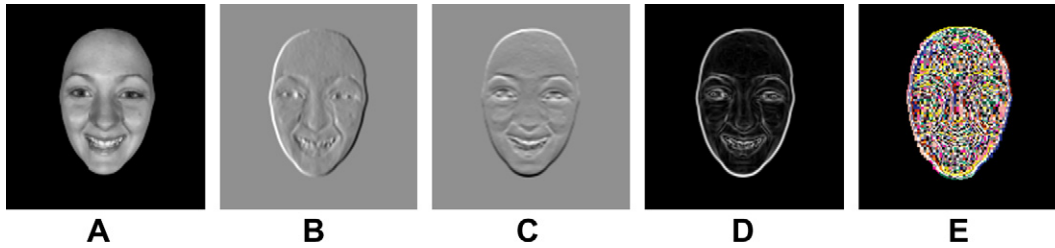


Fig. 4. (A) Original image, (B) horizontal gradient image, (C) vertical gradient image, (D) gradient magnitude image and (E)  $LBP^{\mu 2}$  map of gradient magnitude image.

#### 4.3. Multi-scale LBP

Multi resolution analysis can be achieved by using different values of  $P$  and  $R$ . The  $LBP^{ms}$  has been proven to outperform standard LBPs for face recognition [2] and frontal view facial expression recognition [24]. The first parameter for  $LBP^{ms}$  to be considered is the neighborhood size,  $P$ . A large neighborhood increases the size of the feature histogram and increases the computational cost of each LBP image, while a small neighborhood could result in loss of important information. The second parameter considered is the number of multi-scale operators. A small number of operators might not provide sufficient information for accurate facial expression recognition, where as a large radius operator can decrease the number of uniform patterns which can effect the accuracy of this approach. Here  $LBP^{ms}$  is  $LBP^{\mu 2}(8, R)$ , where  $R = (1, \dots, 8)$  is applied to face images to extract the  $LBP^{ms}$  histogram.

Liao et al. [10] introduces a different approach to multi scale analysis using LBPs. This approach replaces the comparison between single pixels to average gray-values of subregions. Liao et al. argue that this approach is more robust to noise, however it also loses sensitivity and can be distorted by illumination differences.  $LBP^{ms}$  as defined below is more robust because the LBP are calculated for single pixels.

Features calculated over over a local  $3 \times 3$  area cannot encode the larger structures of the face. Thus  $LBP^{ms}$  is more robust, it encodes the micro structures of the face but also the macro structures which provide a more extensive description than the basic LBP operator. Fig. 5 shows the different  $LBP^{\mu 2}$  that contribute to  $LBP^{ms}$ . This figure shows how for smaller  $R$ ,  $LBP^{\mu 2}(8, R)$  captures more detail at the micro scale and for larger values of  $R$  a more structural representation of the face.

$LBP^{ms}$  is formulated in the same way as other features in Sections 4.1 and 4.2. However the final vector will concatenate histograms from each sub-block from eight different  $LBP^{\mu 2}$  maps.

#### 4.4. Local gabor binary patterns

Gabor wavelets have been shown to be suitable for image decomposition and representation when the task is the derivation of local and discriminative features. Gabor filters have been successfully applied to facial expression recognition [11]. Gabor wavelet kernels are similar to the receptive field profiles of the mammalian cortical simple cells. These kernels are popular for vision processing as they display desirable characteristics of spatial locality and orientation selectivity.

The combination of gabor and LBP further enhances the power of the spatial histogram, and exploits multi-resolution and multi-orientation gabor decomposition. LGBP were initially used for face recognition [31]. LGBP are impressively insensitive to appearance variations due to lighting and misalignment [31].

To extract LGBP, the images are convolved with the gabor filters as follows:

$$G(\mu, \nu) = I(x, y) * \psi_{\mu, \nu}(z) \quad (12)$$

where:

$$\psi_{\mu, \nu}(z) = \frac{\|k_{\mu, \nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu, \nu}\|^2 \|z\|^2}{2\sigma^2}} \left[ e^{ik_{\mu, \nu} z} - e^{-\frac{\sigma^2}{2}} \right] \quad (13)$$

$$k_{\mu, \nu} = k_{\nu} e^{i\phi_{\mu}}, \quad k_{\nu} = 2^{-\frac{\nu+2}{2}} \pi, \quad \phi_{\mu} = \mu \frac{\pi}{8} \quad (14)$$

where  $\mu$  and  $\nu$  define the orientation and scale of the gabor filters,  $z = (x, y)$  and  $\|\cdot\|$  denotes the norm operator. Five scales are used

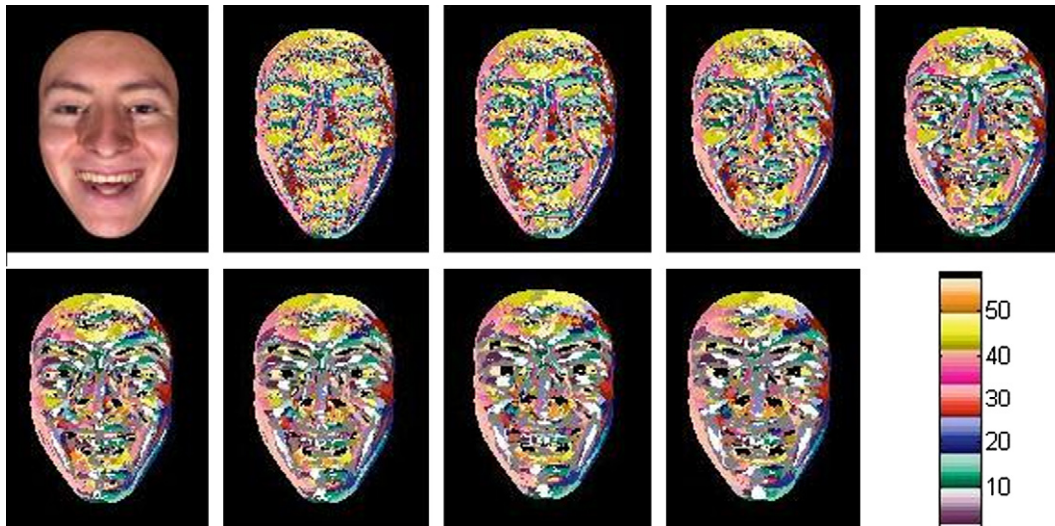


Fig. 5. Different  $LBP^{\mu 2}$  maps used for  $LBP^{ms}$ . Top row left to right, original face image and  $LBP^{\mu 2}(8, 1 \dots 4)$  and on the bottom row  $LBP^{\mu 2}(8, 5 \dots 8)$ .

$v \in \{0, \dots, 4\}$  and eight orientations  $\mu \in \{0, \dots, 7\}$ . These gabor kernels form a bank of 40 different filters.

In order to reduce the dimension of the *LGBP* feature vector,  $LBP^{\mu 2}$  are applied to the gabor maps. The *LGBP* feature vector is formed from 40 gabor magnitude maps, where each map is divided into 64 sub-blocks. The feature vector for the *LGBP* feature is created by concatenating the histograms from each sub-block (similar to Eq. (8)) in each of the gabor magnitude maps. Table 1 summarizes the features formulated in this section.

#### 4.5. Feature extraction

Psychophysical studies in saccadic eye movements [17] indicate that local appearance is important for classification. People can recognize objects when they seek regions where discriminating information is located. *LBP features computed over the whole face represent only the micro patterns without any information about their locations.* Keeping the information about the spatial relationship is very important for facial expression recognition. The approach presented in this paper utilizes this finding by dividing face images into sub blocks and comparing the similarities between these sub-blocks. *The face image is divided into 64 sub-blocks for feature extraction. This is a proven method for accurate facial expression recognition [4,24].* This representation captures local texture and global shape of face images. Fig. 6 shows how the face images are partitioned by a  $8 \times 8$  grid into 64 sub-blocks. Then a histogram of *LBP* features are accumulated and each histogram is concatenated to form a feature vector.

### 5. BU3DFE dataset experiments

Most facial expression databases available have face images of frontal view only. The BU3DFE database [29] provides 3D textured models of six prototypical facial expressions, from which the effects of pose can be investigated by extracting projected 2D images at different yaw angles. In the BU3DFE database, there are 100

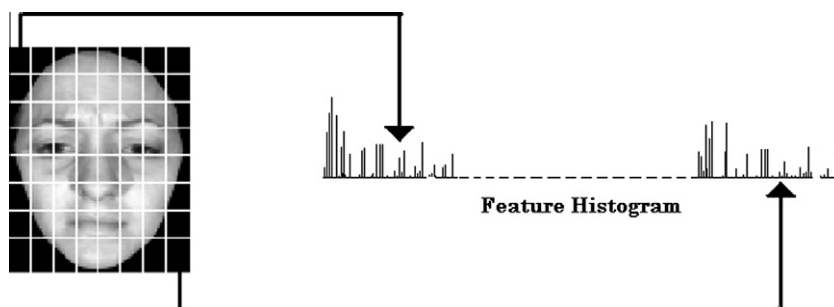
subjects, including undergraduates, graduates and faculty from the State University of New York Binghamton. Age ranges from 18 years to 70 years old. The database consists of 60% female and 40% male with a variety of ethnicity (White, Black, East-Asian, Middle-east Asian, Indian, and Hispanic Latino). Subjects perform the facial expressions in front of a 3D face scanner. Every subject performs each of the six prototypical expressions as well as neutral. Each expression is captured at four different intensity levels (see Fig. 8). Other popular databases do not include different intensities for facial expressions (JAFPE [13] and Cohn-Kanade [8]). This data allows an investigation of how the approach presented in this paper is affected by different intensities of facial expressions. When using the BU3DFE database, images are re-projected from a 3D textured model in OpenGL, resulting in five different poses corresponding to  $0^\circ$ ,  $30^\circ$ ,  $45^\circ$ ,  $60^\circ$  and  $90^\circ$  yaw angles (see Fig. 7). Fig. 7 shows how the face image is divided into sub-blocks and a feature histogram is created from each sub-block. *The final feature vector is a result of concatenating the histograms from each sub-block together.*

#### 5.1. Experiments

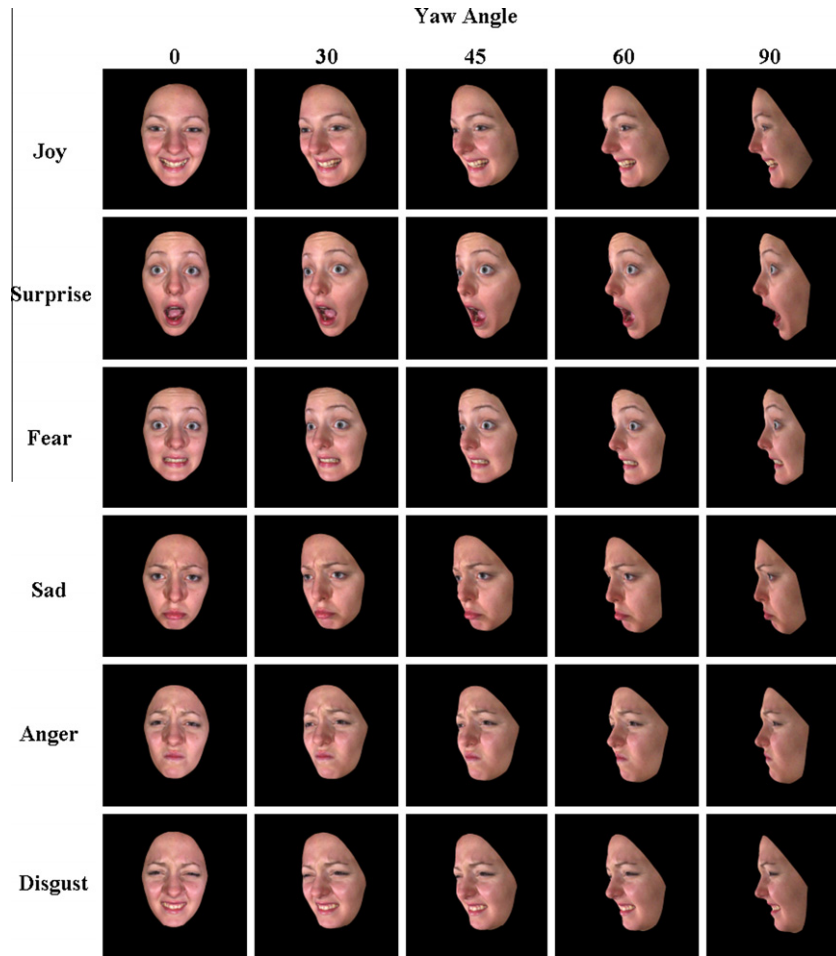
Pose variation typically occurs in human to human interaction by changes in yaw angle. In this section experiments are carried out to classify each of the prototypical expressions at five different yaw angles, this is the same data used in [7] allowing comparison of results. All results on the BU3DFE database are presented using 10-fold cross validation to test the generalization performance of this approach. Training sets of 90 subjects and test sets of 10 subjects were randomly selected, so all expressions for each subject belong in the same group. In total 48,000 images are used for experiments in this section. *In an attempt to classify pose and expression, a sequential approach is used. First a pose classifier is trained on five different views, secondly a view dependent facial expression classifier is trained.* Experiments for pose estimation achieve 100% success rate over the five yaw angles for all features. This is due to the difference in yaw angle being significant

**Table 1**  
Summary of different features used in Section 5.

Feature	Description	Properties	Dimensions
$LBP^{riu2}$	Uniform rotation invariant local binary patterns	Features offer rotation invariance, but poor descriptive abilities	640
$LBP^r$	Rotation invariant local binary patterns	Feature offers rotation invariance, but poor descriptive abilities	2304
$LBP^{gm}$	Uniform local binary patterns obtained from gradient magnitude image	Features encode the magnitude of local variation	3776
$LBP^{\mu 2}$	Standard local uniform binary patterns with a neighborhood of 8 pixels and a radius of 1 pixel	Offers illumination invariance and is computationally efficient	3776
$LBP^{ms}$	Multi-scale local binary patterns where radius varies from 1 to 8 pixels	Multi-scale analysis can encode the micro features of the face plus features at the structural level	30,208
<i>LGBP</i>	Local binary patterns are extracted from gabor images, where 40 different gabor images are composed from applying gabor kernels at different scales and orientations	Gabor filters offer strong illumination invariance as well as powerful descriptive features. However, the feature vector has high dimensionality	151,040



**Fig. 6.** Feature vectors are built by concatenating feature histograms from each sub-block of the grid.



**Fig. 7.** Data from BU3DFE database showing the different yaw angles used for each facial expression. The textured 3D models were re-projecting at different yaw angles to create the 2D images.

(between  $15^\circ$  and  $30^\circ$ ) and the relatively clean nature of the synthetic data. In Section 6 we address this issue by performing similar experiments using live data. However, the aim of these experiments are primarily to look at the effect of pose and features type on recognition accuracy.

#### 5.1.1. Effects of different features and resolution

Table 2 shows the overall recognition results for the features formulated in Section 4 over four different resolutions. Bold values in tables throughout this paper, highlight the highest value for a particular column. Interestingly, there is no significant performance increase for higher resolutions, as in general it is the faces micro features which represent deformation. **Less than 3% difference in performance for features over the four different resolutions. This shows the power of LBP features to capture important information for facial expression recognition at low resolutions.**

**Table 2**

Overall performance of features for four different resolutions.

	$32 \times 44$	$44 \times 62$	$64 \times 88$	$80 \times 110$
$LBP^{riu2}$	<b>47.28</b>	46.12	46.31	46.32
$LBP^{ri}$	<b>47.53</b>	46.28	45.93	46.56
$LBP^{gm}$	52.91	51.49	53.2	<b>53.29</b>
$LBP^{u2}$	<b>58.44</b>	57.33	57.12	56.24
$LBP^{ms}$	62.41	62.9	64.98	<b>65.02</b>
LGBP	66.76	67.84	<b>67.96</b>	66.79

Features  $LBP^{ri}$  and  $LBP^{riu2}$  perform poorly on facial expressions. This is most likely because the histograms are not descriptive enough to disambiguate facial expressions correctly. Interestingly, over all resolutions there is less than 1% difference in performance between these features. Thus proving that the uniform patterns for features  $LBP^{riu2}$  provide as much discrimination ability as  $LBP^{ri}$  features.

Another interesting observation is  $LBP^{gm}$  performed worse than  $LBP^{u2}$ . Thus the derivative based  $LBP^{gm}$ , which encodes velocity of local variation is outperformed by the standard  $LBP^{u2}$  on raw image data for classifying facial expressions.  $LBP^{ms}$  outperforms standard  $LBP^{u2}$  by up to 8%, utilizing the multi scale analysis. This result is interesting as it highlights the importance of multi-scale analysis for facial expression recognition. **LGBP outperforms all other features because of multi-resolution analysis combined with multi-orientation analysis.** Although  $LBP^{ms}$  combines multi-resolution analysis with multi-orientation analysis, the gabor representation proves more powerful as a texture descriptor. However the LGBP representation is more computationally expensive.

Tables 3–8 show the confusion matrices for features formulated in Section 4. Some general trends can be observed that are common to all features. The best performing expression is *surprise* followed by *joy*. The high recognition rate of expression *surprise* and *joy* can be attributed to the large amount of deformation of the face for these expressions. *Fear* consistently has the lowest recognition rate. A contributing factor to the poor performance of the *fear* expression is its confusion with the *joy* expression. Similar



**Table 3**  
Confusion matrix for *LGBP*.

Feature	Anger	Disgust	Fear	Joy	Sadness	Surprise
Anger	<b>63.06</b>	8.81	3.50	1.88	19.62	3.13
Disgust	14.75	<b>63.25</b>	6.63	5.75	6.44	3.19
Fear	6.12	9.38	<b>50.94</b>	14.06	10.19	9.31
Joy	2.56	4.81	10.37	<b>79</b>	1.69	1.56
Sadness	17.56	2.81	5.88	1.44	<b>68.13</b>	4.19
Surprise	1.31	4.69	5.50	1.56	3.56	<b>83.37</b>

**Table 4**  
Confusion matrix for *LBP<sup>ms</sup>*.

Feature	Anger	Disgust	Fear	Joy	Sadness	Surprise
Anger	<b>55.31</b>	15.31	4.94	1.31	19.87	3.25
Disgust	12	<b>63.31</b>	7.06	4.50	7.56	5.56
Fear	6.50	9.25	<b>49</b>	12.19	11.06	12
Joy	3.37	6.25	9.31	<b>76.94</b>	1.06	3.06
Sadness	15.75	7.37	6.31	3.13	<b>63.38</b>	4.06
Surprise	2.81	5.63	3.38	2.50	3.50	<b>82.19</b>

**Table 5**  
Confusion matrix for *LBP<sup>u2</sup>*.

Feature	Anger	Disgust	Fear	Joy	Sadness	Surprise
Anger	<b>30.12</b>	16.56	1.81	8.93	27.50	15.06
Disgust	5.87	<b>55.75</b>	2.37	13.43	8.68	13.87
Fear	3.75	10.25	<b>18.50</b>	30.56	12.50	24.43
Joy	2.37	2.06	1.87	<b>84.31</b>	1.56	7.81
Sadness	10.25	7.00	2.37	6.25	<b>60.56</b>	13.56
Surprise	1.75	2.43	0.62	3.50	3.50	<b>88.18</b>

**Table 6**  
Confusion matrix for *LBP<sup>gm</sup>*.

Feature	Anger	Disgust	Fear	Joy	Sadness	Surprise
Anger	<b>30.31</b>	12.50	2.12	10.93	29.93	14.18
Disgust	5.81	<b>39.31</b>	3.62	19.87	11.00	20.37
Fear	4.75	9.25	<b>16.62</b>	28.25	13.81	27.31
Joy	2.3125	1.93	2.75	<b>81.31</b>	2.62	9.06
Sadness	8.06	4.56	1.93	6.31	<b>64.50</b>	14.62
Surprise	1.68	1.81	1.43	3.12	4.25	<b>87.68</b>

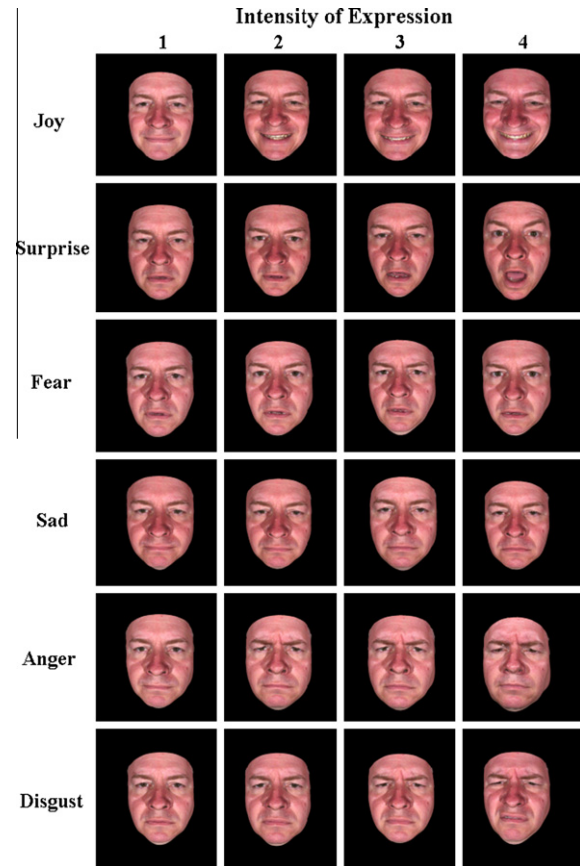
**Table 7**  
Confusion matrix for *LBP<sup>ri</sup>*.

Feature	Anger	Disgust	Fear	Joy	Sadness	Surprise
Anger	<b>23.44</b>	11.81	3.81	11.06	31.62	18.25
Disgust	9.31	<b>29.87</b>	3.68	23.62	11.68	21.81
Fear	5.25	8.68	<b>13.18</b>	34.62	16.43	21.81
Joy	2.87	4.50	3.31	<b>73.87</b>	3.18	12.25
Sadness	8.75	3.93	5.56	10.12	<b>58.12</b>	13.50
Surprise	2.62	3.18	1.62	7.06	4.62	<b>80.87</b>

**Table 8**  
Confusion matrix for *LBP<sup>riu2</sup>*.

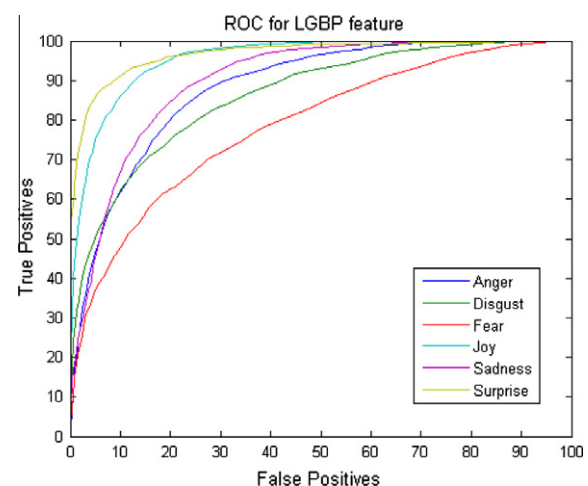
Feature	Anger	Disgust	Fear	Joy	Sadness	Surprise
Anger	<b>21.31</b>	13.12	4.06	11.00	31.87	18.62
Disgust	9.00	<b>30.18</b>	3.93	23.43	11.31	22.12
Fear	4.87	9.06	<b>13.37</b>	34.87	15.87	21.93
Joy	2.43	3.93	4.12	<b>73.81</b>	3.50	12.18
Sadness	9.12	4.00	5.18	9.62	<b>58.56</b>	13.50
Surprise	2.68	3.75	1.43	6.87	4.56	<b>80.68</b>

deformation around the mouth occurs for both expressions. Another common result for all features is the high confusion between the expressions *anger* and *sadness*. These two expressions have the



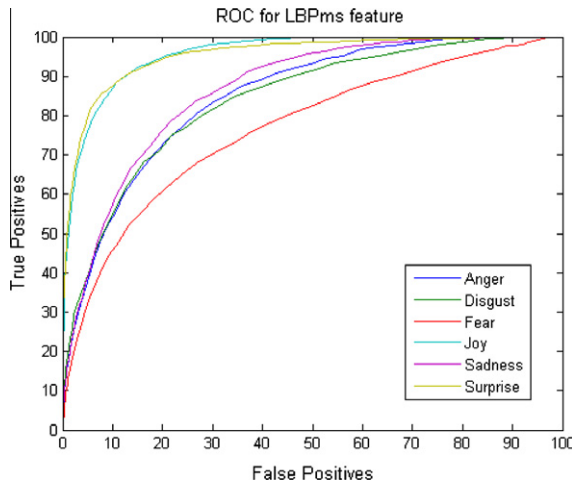
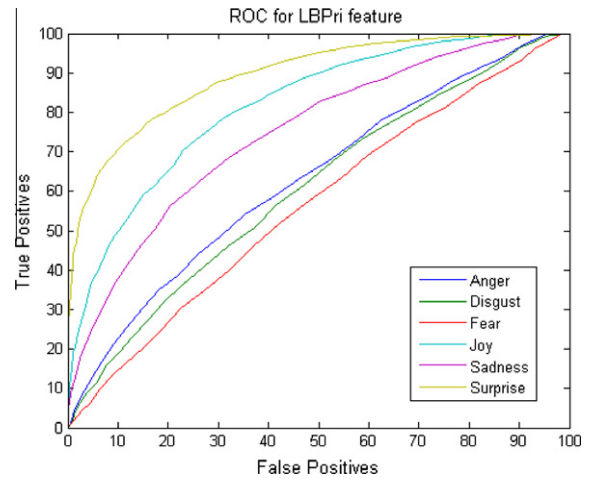
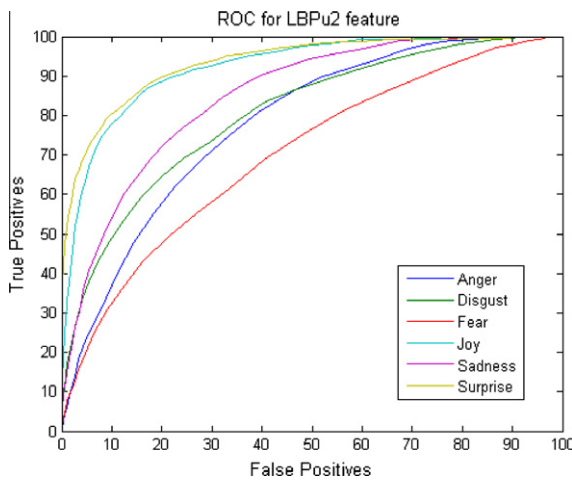
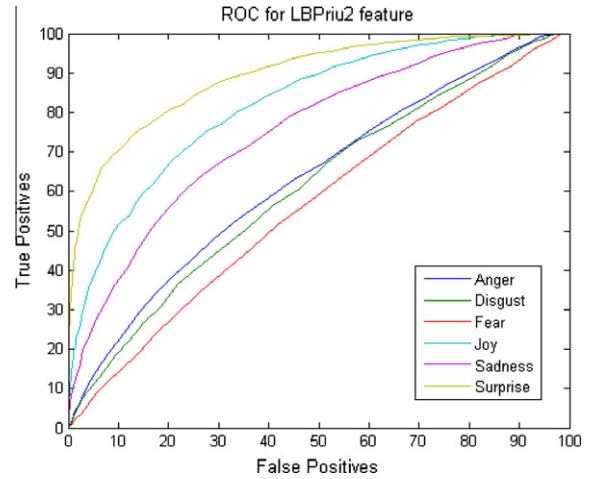
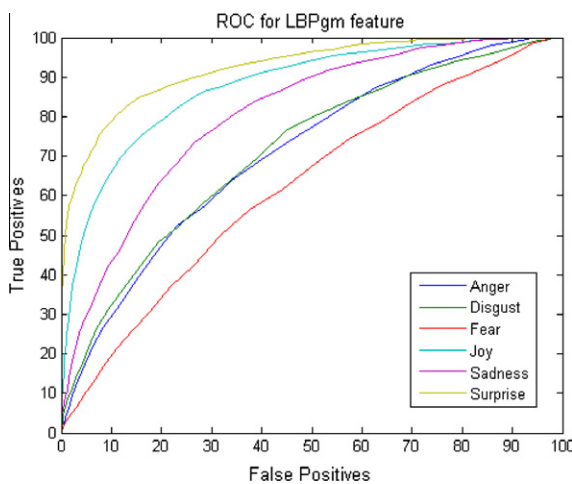
**Fig. 8.** Examples of the different intensities for each facial expression in the BU3DFE database.

least amount of facial movement and thus are difficult to distinguish from each other. From Fig. 8, it can be seen how difficult it is to distinguish the expressions *sadness* and *anger* particularly for the lower intensities. This is a common problem in facial expression recognition as both expression are subtle and hard to distinguish. Figs. 9–14 show the ROC curves for the different features formulated in Section 5.1. The same trends as observed from the confusion matrices are evident from the ROC curves. *Surprise* and *joy* are consistently the best performing expressions, while *fear* performs poorest.



**Fig. 9.** ROC curves for *LGBP* feature.



Fig. 10. ROC curves for  $LBP^{ms}$  feature.Fig. 13. ROC curves for  $LBP^{ri}$  feature.Fig. 11. ROC curves for  $LBP^{u2}$  feature.Fig. 14. ROC curves for  $LBP^{riu2}$  feature.Fig. 12. ROC curves for  $LBP^{gm}$  feature.

Tables 7 and 8 for features  $LBP^{ri}$  and  $LBP^{riu2}$  have very similar results. The performance of expressions *surprise* and *joy* are high, around 74% and 81% respectively. These results are impressive given the reduced capacity of the histograms. However results

for expressions *anger*, *disgust* and *fear* are very poor, with particular poor results (13%) for the *fear* expression.

Focusing on Tables 7 and 5, the influence of orientation analysis for facial expression recognition can be quantified. From Table 5, feature  $LBP^{u2}$  outperforms feature  $LBP^{ri}$  by about 10%. Looking at the confusion matrix for both features, it can be observed that the influence of orientation analysis effects all expression with the largest performance increase for expression *disgust*. This could be attributed to the importance of the orientation of the eyebrows for the *disgust* expression. The recognition rate for *sadness* only increases by about 5%. This indicates that the *sadness* expression does not rely on orientation analysis as much as other expressions.

Tables 3 and 4 show confusion matrices for the best performing features,  $LGBP$  and  $LBP^{ms}$  respectively.  $LGBP$  outperform  $LBP^{ms}$  for all expressions except *disgust*, where results are similar. The largest confusion occurs between expressions *anger* and *sadness* for both sets of features. Confusion for expressions *disgust* and *anger* is also evident in both Tables 3 and 4.

### 5.1.2. Effects of pose

In the following section an investigate of which pose is optimal for facial expression recognition and how pose variations effects particular facial expressions is presented. Fig. 15 shows the overall recognition rate for each yaw angle for each feature and resolution.

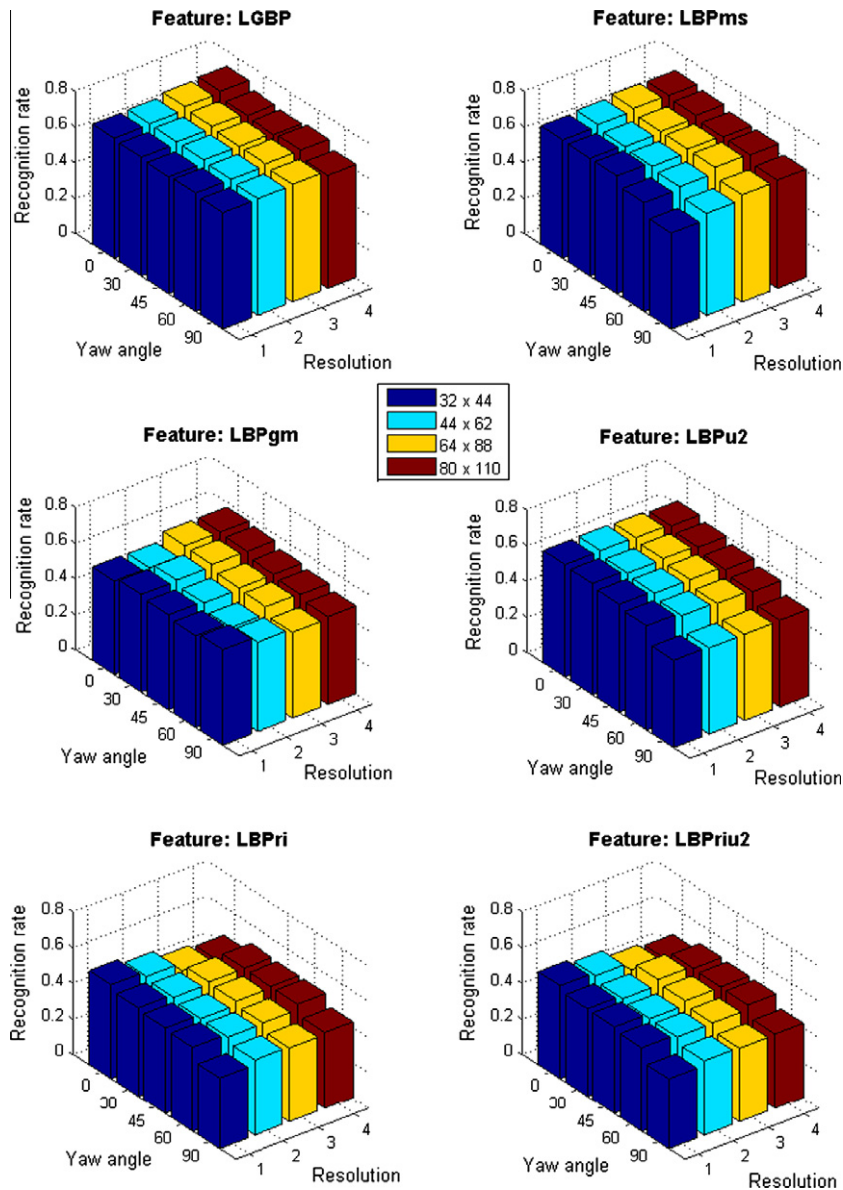


Fig. 15. Recognition rate of view independent classifiers for all expressions, for all features presented in Section 4.

Frontal pose is the optimal view over all resolutions for features *LGBP*, *LBP<sup>ms</sup>* and *LBP<sup>u2</sup>*. These features are the 3 best performing features for facial expression classification (see Table 2). However, from Fig. 15, it is also evident that performance does not decrease significantly due to yaw variation. In particular *LGBP* achieved impressive results for large yaw angles. The performance for *LGBP* from frontal to profile is relatively uniform. Where as *LBP<sup>ms</sup>* performance drops significantly for yaw angles 60° and 90°. In general this trend is also evident for features *LBP<sup>u2</sup>* and *LBP<sup>gm</sup>*. Fig. 15 also shows that weaker features, in particular *LBP<sup>ri</sup>* and *LBP<sup>riu2</sup>*, sometimes perform better at non-frontal views. But even in this scenario, the optimal yaw angle varies. This provides evidence that selection of features, plays an important role in answering the question which view is optimal for facial expression recognition. Weaker features might not be efficient enough to utilize the discriminatory information available at frontal pose.

Another important question is how does yaw variation effect individual expression recognition performance. Fig. 16 shows the performance of each expression over five yaw angles for *LBP<sup>u2</sup>*, *LBP<sup>ms</sup>* and *LGBP* over four resolutions. It does not follow that

because frontal view is optimal for overall expression recognition, that individual expressions are optimal at frontal view. This is confirmed by Fig. 16. *Sadness* performs remarkably well at profile view (yaw 90) over all three features, often outperforming other views. For the *LGBP* feature over all four resolutions, *sadness* is consistently classified best at non-frontal view. This is most likely due to the lip movement which protrudes from the face for the expression and is more evident at large yaw angles. *Anger* is also classified best at non-frontal view for *LGBP* but not for other features. Another interesting finding is the performance drop of the expression *joy* as the yaw angles increases for the *LGBP* feature. This suggests that important discriminatory information is lost as the yaw angle increases for the *joy* expression. This finding is only evident for *LGBP* and not the other features, suggesting that complementary information between different features exists. Also, from these results it is clear that *LBP<sup>u2</sup>* suffers because of its inability to classify the expressions of *anger* and *fear* particularly at large yaw angles. For features *LBP* and *LBP<sup>ms</sup>*, performance generally degrades with larger yaw angles with the exception of the *sadness* expression. This is also true for *LGBP*, with the exception of *anger*, which is

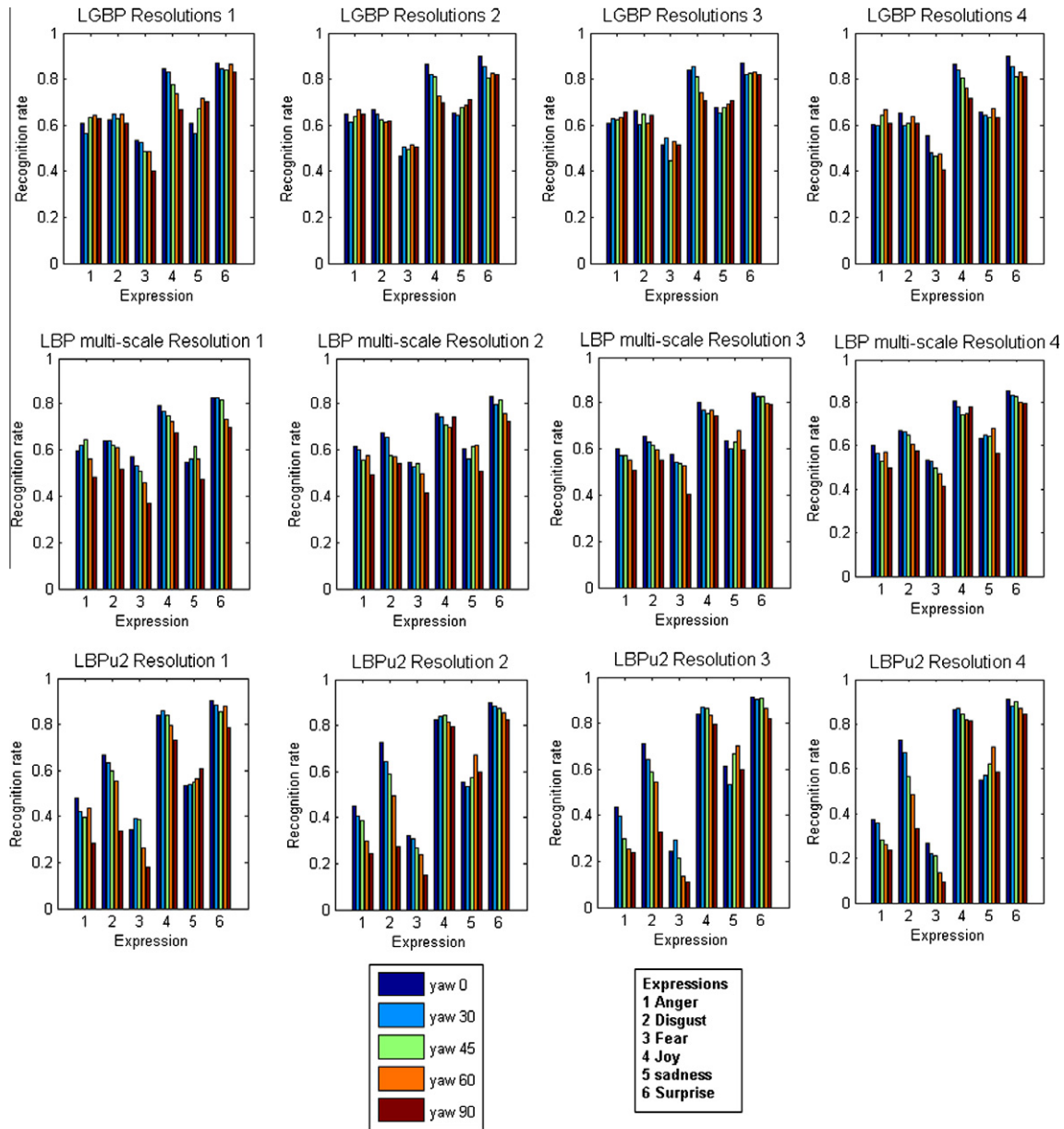


Fig. 16. Performance of individual expressions for each yaw angle.

recognized best at non-frontal view. This suggests that *LGBP* is exploiting different multi-orientated filters at different angles. Observations in this section are drawn from Fig. 16 which appear for the 4 different resolutions.

## 5.2. Discussion

Previous studies employing geometric features on the BU3DFE database are [6,7,32]. Hu et al. [7] presented evidence that non-frontal views are best for automatic recognition of facial expressions over varying yaw angles. Geometric points around the salient features of the face are used. The 2D displacement of each feature point for an expression against the same point for the neutral expression of that subject is calculated and normalized. Utilizing an SVM for classification [7] suggested that 45° yaw angle performed the best. Hu et al. [6] also report results for similar experiments. Features are extracted from an area around the feature points using SIFT, HoG and *LBP* features. Results show optimal performance at 30°. Zheng et al. [32] using the same experiments

as above using similar features to [6] show best performance at 60°. The main conclusion of [7,6,32] is that non-frontal views are better than frontal view for a computer to recognize facial expressions. However, as can be seen, Fig. 15 shows conflicting results. This could be attributed to the different type of features used. The disparity between results could indicate that features play an important part in answering the question of which view is optimal.

Results using *LGBP* show that actual performance is relatively consistent across yaw angles. Also from other research mentioned in this section the performance difference across yaw angles is marginable [7,6,32]. Head pose recognition achieved 100% in experiments. The problem of head pose classification is simplified by the cropped 3D models and the large interval in yaw angles (up to 30°, see Fig. 7).

The BU3DFE database captured each facial expression at four different intensities. Other popular facial expression databases do not include such variety. This allows analysis of how intensity effects the performance of the methods. Table 9 shows the performance of features for the different intensities of expressions. As

**Table 9**  
Performance of features for four different intensities.

Features	Intensity			
	1	2	3	4
$LBP^{riu2}$	38.33	46.67	48.50	51.79
$LBP^i$	38.96	46.54	48.71	52.04
$LBP^g$	44.58	53.21	56.75	58.63
$LBP^{u2}$	46.08	57.46	58.75	62.67
$LBP^{ms}$	53.92	64.46	69.25	72.46
$LGBP$	<b>56.83</b>	<b>68.63</b>	<b>73.04</b>	<b>77.67</b>

expected, higher intensity leads to higher recognition rates.  $LGBP$  performs consistently best with intensity level 4 achieving an average recognition rate of 77.67%. Recognition rates of over 90% have been achieved using similar approaches to this paper for the Cohn-Kanade database. But these results are for exaggerated expressions only. This database offers a more realistic evaluation because of the challenge of facial expression recognition at different intensities.

Another interesting observation from results presented in Section 5.1.1 is the performance of uniform patterns for facial expression recognition. The use of uniform patterns has been justified by the results. Results show no significant difference between features  $LBP^i$  and  $LBP^{iu2}$ . Similar observations were found for  $LBP$  and  $LBP^{u2}$  for frontal facial expression recognition [24]. The advantage of using uniform patterns is the reduction in the size of the feature histogram without significant losses of accuracy.

It is evident from Fig. 16 that complimentary information is present in both  $LGBP$  and  $LBP^{ms}$  due to different performance at different yaw angles. Combining the feature vectors together as input to an SVM, allows us to capture the performance of both features. An overall performance of 71.1% was achieved for a combined feature vector of  $LGBP$  and  $LBP^{ms}$  and gives a performance increase of 3% over the  $LGBP$  features. Table 10 shows a comparison of geometric and appearance feature based approaches. All approaches use an SVM as the classifier and are tested on similar yaw variations. However the geometric based method [7] requires manually labeled feature points of the mouth, eyes and eyebrows.

The best performing features were  $LGBP$  and  $LBP^{ms}$  on the BU3DFE dataset. These features performed best at frontal view. However, as can be seen from Fig. 15 performance is relatively consistent across pose. Observing the overall results in Table 2 higher recognition results correspond to the complexity of feature histograms. In the experiments presented in this section synthetic data was used. The face models were cropped (as can be seen in Fig. 7) and because images are re-projected from a 3D textured model, there is less variability in the synthesized dataset. The next section introduces experiments on a live dataset.

## 6. Multi-pie dataset experiments

This section is going to build on the previous section by applying the best features and validating results on real data. To do this we will use the multi-pie database to evaluate the performance of the approach presented in this paper. The multi-pie database [5]

**Table 10**  
Comparison of features methods.

Feature method	Results
$LBP^{ms}$	65.02
Geometric based [7]	66.5
$LGBP$	67.96
$LGBP/LBP^{ms}$	71.10

contains images from 337 subjects. Subjects are predominantly male (70%). 60% of subjects were European Americans, 35% Asian and 3% African Americans. The average age of the subjects was 28 years old. Data was captured during four sessions over a six month period. In each session, subjects were instructed to display various facial expressions (*neutral, smile, surprise, squint, disgust* and *scream*, see Fig. 17). Before each session subjects were shown examples of the particular facial expression from the Cohn-Kanade database [8]. Thirteen cameras were located at head height in 15° intervals. 100 subjects were selected so all subjects were present at all four recording sessions.

### 6.1. Face detection

The Viola and Jones face detector [26] is used to extract the face region for all poses. This is available from the OpenCV library [21]. The face detector uses boosted cascades of harr-like features. The frontal detector was used for poses 0°, 15° and 30°. The profile cascade was used for poses 45°, 60°, 75° and 90°. The performance of the frontal detector was superior to the profile detector. Some false positives and missed detections were observed for the profile detector mainly at 75° and 90°. Most of the missed detection occur when part of the face is occluded by facial hair. Images that were incorrectly classified were corrected and labeled manually. False positives were removed manually. Fig. 18 shows the results of using the opencv frontal and profile detectors on the multi-pie database.

### 6.2. Features

Best practices are taken from experiments presented in Section 4. The best performing features were  $LGBP$  and  $LBP^{ms}$ . Only a difference of around 3% in recognition rate was observed with  $LGBP$  (the better performing feature). These features are applied to the multi-pie dataset. Experiments presented in Section 5 used 3D models which were centered so faces images were perfectly aligned. In this section a more automated approach is presented. However, the inherent noise in the face detector will test the ability of these



**Fig. 17.** Example of facial expressions from multi-pie database. Top row – neutral and smile. Middle row – surprise and squint. Bottom row – disgust and scream.





**Fig. 18.** Opencv frontal and profile face detector results. The frontal face detector was used for poses 0°, 15° and 30°, while the profile detector was used for poses 45°, 60°, 75° and 90°.

features to discriminate facial expressions in the presence of misalignment errors.

### 6.3. Experiments

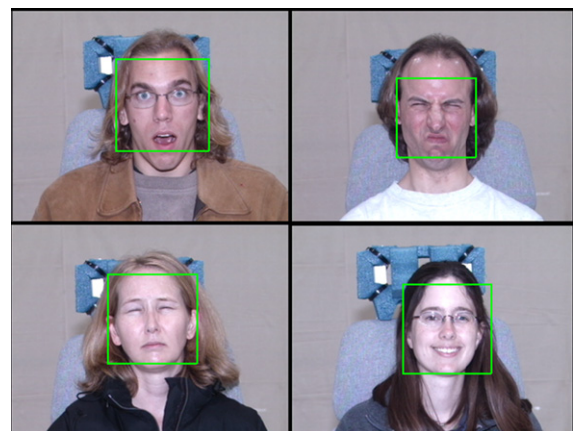
For the experiments presented in this section, seven different poses (0°, 15°, 30°, 45°, 60°, 75° and 90° yaw angles), are considered (see Fig. 18). 100 subjects were selected so that all subjects were present at all four sessions and thus for each subject all expressions were available. In total 4200 images were used for experiments. Images were resized to  $320 \times 240$ , where the typical face detection size was around  $100 \times 100$  pixels. To test the algorithms generalization performance, all experiments in the following sections are based on 10-fold cross validation. Training and test sets were divided 80–20%. This 20% testing data is taken from subjects that were not present in the training data. This ensures that any features extracted for classification provide person independent facial expression recognition. Features are extracted using a grid with 64 sub-blocks (similar to Fig. 6).

#### 6.3.1. Head pose and expression classification

To classify pose and expression, a cascade approach is adopted where the classification task is divided into two steps. First, a pose classifier is trained over seven views from frontal to profile view in 15° increments. Secondly, a pose dependent expression classifier is trained to classify expressions. When training the pose classifiers all expressions for each pose are including in the training sets. Thus the difference between expressions is regarded as within-class variance. Expression classifiers are trained for each pose. In total 42

expression classifiers are trained (seven different poses with six pose specific expression classifiers at each pose). A multi-class SVM is used for final classification (one against all approach).

From Fig. 18 it can be seen that at some poses part of the background, including parts of the chair used in the multi-pie recording is present in the area returned by the face detectors. This is most evident at frontal view. However, Fig. 19 shows other subjects where the background is different due to clothes, hair and position of subjects head. Thus, enough variability exists in the



**Fig. 19.** Examples of variations present in the multi-pie database. The area returned by the face detector is subject dependent and does not always contain background features.

**Table 11**  
Recognition rates for head pose and overall facial expression recognition on multi-pie database.

Features	Pose	Expressions
<i>LBP<sup>ms</sup></i>	99.13	73.98
LGBP	<b>99.45</b>	<b>80.17</b>

**Table 12**  
Facial expression recognition results for each yaw angle.

	0°	15°	30°	45°	60°	75°	90°
<i>LBP<sup>ms</sup></i>	76.7	<b>80.5</b>	70.3	69	78.6	63	73.8
LGBP	82.1	<b>87.3</b>	75.6	77.8	85	71	75.9

**Table 13**  
Confusion matrix for facial expressions over all yaw angles for *LBP<sup>ms</sup>* features.

	Neu	Smi	Sur	Squ	Dis	Scr
Neu	<b>73.92</b>	11.57	2.98	8.91	3.41	0.66
Smi	9.21	<b>78.04</b>	4.04	4.79	3.62	1.74
Sur	3.41	3.40	<b>81.01</b>	2.54	1.89	9.21
Squ	9.28	8.84	2.90	<b>60.11</b>	18.71	1.60
Dis	5.51	4.85	1.74	14.87	<b>69.21</b>	5.27
Scr	0.15	1.15	12.95	0.94	3.48	<b>81.57</b>

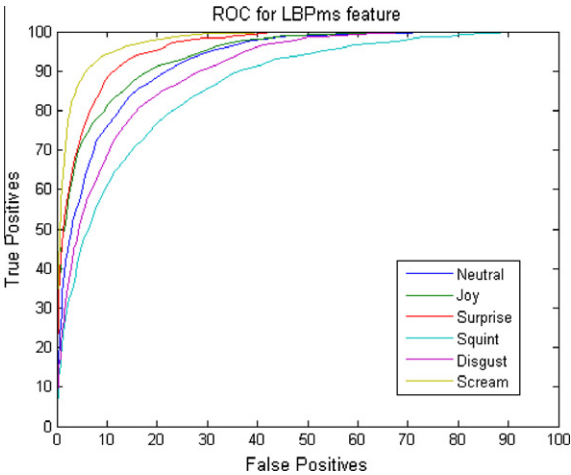
**Table 14**  
Confusion matrix for facial expressions over all yaw angles for LGBP features.

	Neu	Smi	Sur	Squ	Dis	Scr
Neu	<b>80.55</b>	8.02	2.75	6.87	2.67	0.58
Smi	7.54	<b>82.74</b>	2.61	5.07	2.62	0.87
Sur	1.03	3.55	<b>88.67</b>	0.87	1.81	5.52
Squ	8.61	7.45	1.37	<b>66.26</b>	16.89	0.87
Dis	4.12	3.55	1.02	14.70	<b>74.81</b>	3.25
Scr	0.14	0.94	8.52	0.36	2.18	<b>88</b>

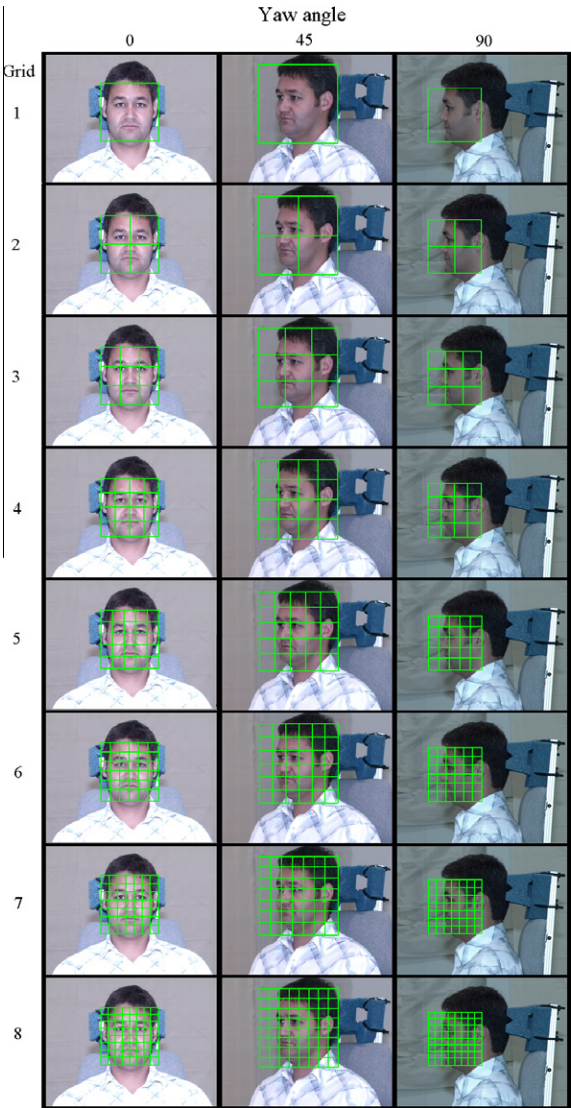
dataset to suggest that background features are not a contributing factor to training the head pose classifiers. Given that view specific classifiers were trained, the background is the same for all expression and will not effect the training of facial expression classifiers.

6.3.2. Results

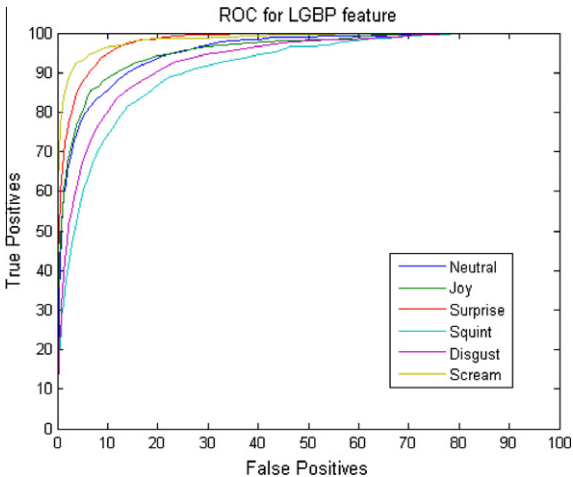
Table 11 shows the overall results for head pose and facial expression classification. As expected, LGBP outperformed *LBP<sup>ms</sup>* for both head pose and facial expressions. Both features achieved head pose



**Fig. 21.** ROC curves for *LBP<sup>ms</sup>* feature.



**Fig. 22.** Examples of different grids applied to the face region to extract histograms of features.



**Fig. 20.** ROC curves for LGBP feature.

recognition results of over 99% averaged over seven poses. LGBP significantly outperforms *LBP<sup>ms</sup>* by over 6% for facial expressions

**Table 15**

Table showing results for pose estimation averaged for all poses for different sampling grids. Where grid 1 is a global histogram and grid 8 being a  $8 \times 8$  grid over the face (see Fig. 22).

	Grid size							
	1	2	3	4	5	6	7	8
Multi-scale LBP	96.94	98.64	99.05	99.30	99.03	99.01	99.09	99.13
LGBP	98.65	99.17	99.51	99.45	99.60	99.52	99.48	99.45

over all poses. This performance difference is further evidence for the capabilities of the multi-orientation multi-resolution analysis present in the LGBP features. Table 12 shows results of both features for each head pose. Surprisingly, results at angles  $15^\circ$  and  $60^\circ$  outperform frontal view. View  $15^\circ$  achieves the best results for both features. Another interesting finding for both features is profile views outperform other views, where more of the face is visible. Given some of the problems with occlusion for profile view (discussed in the Section 6.5), this result is interesting.

Tables 13 and 14 show the confusion matrices for  $LBP^{ms}$  and LGBP respectively. In general, the same patterns can be seen for both sets of features. The most confusion occurs between expressions *squint* and *disgust*, due to the expressions having similar deformation around the eyes. In fact, the *squint* expression has some confusion with other expressions including *neutral* and *smile*. This is most likely due to the fact that *squint* is a relatively subtle expression and thus is hard to disambiguate between other expressions. Subtle expressions are hard to distinguish because of the variability across subjects. More confusion is present between expressions *scream* and *surprise*, this can be attributed to the similar deformation of the mouth. *Surprise* usually is associated with raised eyebrows, but for some subjects in this database no noticeable deformation occurs around the eyebrows. This could contribute to the confusion. Figs. 20 and 21 show the ROC curves for  $LBP^{ms}$  and LGBP respectively. The best performing expressions for both features are *surprise* and *scream*. These expressions have lots of deformation and thus are easier to distinguish than more subtle

expressions. Recognition results for the *squint* expression are poor for both features due to the subtle nature of the expression.

#### 6.4. Local versus global feature representation

To investigate the arbitrary nature of using  $8 \times 8$  sub regions, different grid sizes are tested from a global representation ( $1 \times 1$ ) up to a  $64 (8 \times 8)$  sub-block representation. Having a large number of sub blocks can degrade the accuracy in the present of localization errors and also increases the computation cost. A small number of sub-blocks increase the loss of spatial information. For the experiments in this section, eight different grids are applied to sample feature histograms. Fig. 22 shows these grids for three poses.

Table 15 shows the overall performance of  $LBP^{ms}$  and LGBP for head pose recognition over the seven different yaw angles. The

**Table 16**

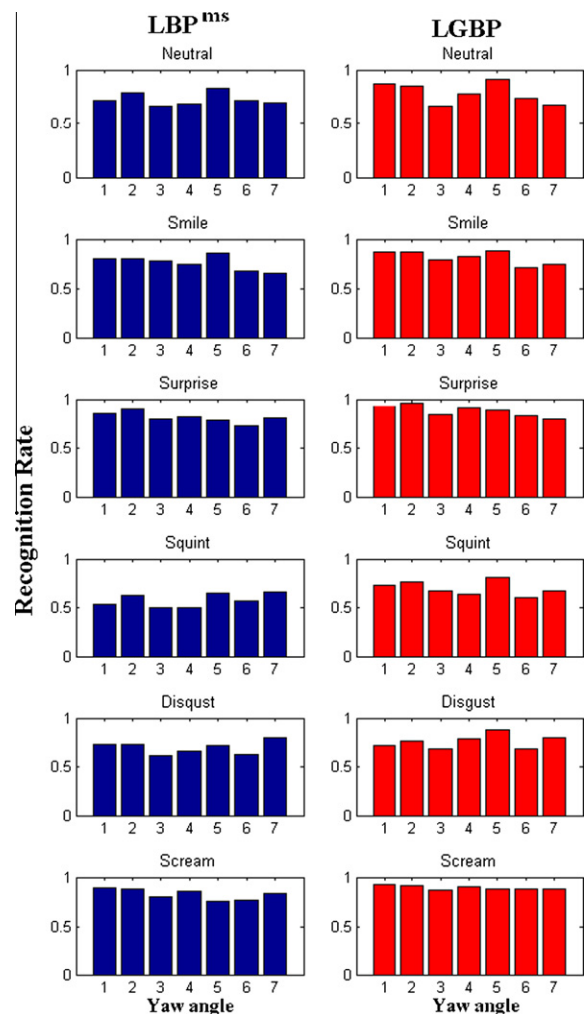
Facial expression recognition results for  $LBP^{ms}$  features on pose dependent data, Rows correspond to different sampling grids and columns correspond to different poses.

Grid size	Pose							Avg.
	0	15	30	45	60	75	90	
1	57.58	60.75	54.17	50.42	61.25	52.75	63.00	57.13
2	66.58	73.75	65.08	67.08	72.25	61.17	73.92	68.55
3	73.33	76.67	68.50	67.67	77.50	66.25	75.33	72.18
4	73.00	77.83	<b>71.42</b>	70.00	<b>78.67</b>	66.67	74.83	73.20
5	73.75	79.00	68.42	70.75	76.58	65.58	<b>76.33</b>	72.92
6	76.37	79.45	69.62	<b>71.62</b>	77.37	<b>68.53</b>	74.78	<b>73.26</b>
7	76.42	79.92	68.33	69.08	78.33	66.25	73.33	73.09
8	<b>76.87</b>	<b>80.90</b>	71.83	69.70	79.78	64.20	73.75	73.86

**Table 17**

Facial expression recognition results for LGBP features on pose dependent data, rows correspond to different sampling grids and columns correspond to different poses.

Grid size	Pose							Avg.
	0	15	30	45	60	75	90	
1	60.75	63.50	57.50	51.58	61.83	51.50	66.50	59.02
2	72.25	76.42	69.83	64.58	73.92	62.00	74.25	70.46
3	80.08	82.08	73.00	72.58	80.92	68.67	77.75	76.44
4	78.83	82.08	75.92	76.83	84.92	71.92	<b>79.58</b>	78.58
5	81.33	83.33	<b>76.67</b>	79.33	87.17	72.17	75.33	79.33
6	<b>84.00</b>	85.50	75.25	<b>80.33</b>	87.58	<b>74.00</b>	76.17	<b>80.40</b>
7	82.08	85.67	76.33	78.83	<b>89.08</b>	73.08	76.50	80.22
8	82.58	<b>87.85</b>	76.58	78.95	85.00	72.50	77.92	80.19



**Fig. 23.** Performance of individual facial expressions for different yaw angles for features  $LBP^{ms}$  and LGBP.



performance of both features is consistently over 99% for most grid sizes. These results indicate the complexity of head pose estimation is significantly reduced by only classifying head pose at 15° intervals. Also the results show that there is no significant performance difference for a global or a local histogram approach. This reflects the nature of the problem as head pose estimation unlike facial expression recognition does not discriminate based on local variations of the face but at a structural level.

Tables 16 and 17 show the results for each yaw variation for the different grids used to extract the feature vectors. In general the trend is the finer the grid the more accurate the results. The lack of spatial information for the global approach reduces the accuracy. For grid sizes 1–3 the difference between  $LBP^{ms}$  and  $LGBP$  is insignificant but as the number of sub-blocks increases the performance of  $LGBP$  is up to 7% greater. For both features the best performance was for a grid size of six (36 different sub-blocks). In general for both features, head poses 15° and 60° perform the best. The highest results for any particular yaw angle was achieved at 60° for grid size 7, with a very high recognition rate of over 89% for the  $LGBP$  feature.

Fig. 23 shows the individual performance of each facial expression for each yaw angle. Comparing the features  $LBP^{ms}$  and  $LGBP$ , Fig. 23 shows that  $LGBP$  outperforms  $LBP^{ms}$  for all expression, but squint performs significantly better for  $LGBP$  than  $LBP^{ms}$ . Once again showing the importance of Gabor filter for the more subtle expressions. The scream expressions performance is particular high and consistent across different poses. This is evident in particular for the  $LGBP$  features, this could be explained by the exaggerated nature of the scream expression in the database. Squint and disgust expressions perform surprisingly well at profile view and

outperform other views. However as with experiments on the BU3DFE dataset performance is relatively consistent over pose.

### 6.5. Discussions

The multi-pie dataset is a very challenging dataset. Figs. 24 and 1 show examples of some of these difficulties. Of the 100 subjects used for experiments, 49 subjects wore glasses in some or all recording sessions. Implications of this are significant given that at different head poses glasses can occlude parts of the eyes and eyebrows, where subtle information can be lost. Other challenging aspects of this dataset are hair covering the eyes and eyebrows for some views. Fig. 24 shows how hair can occlude facial features for views 0°, 75° and 90° amongst others. This database also has subjects with beards and mustaches. Other popular databases including JAFFE [13] and Cohn-Kanade database [8] do not have the same level of variability. Another variable to consider when evaluating the above results is the noise included by the face detector. This noise is more apparent for the profile detector, particularly at 75° and 90° views. Some missed detections occurs at these poses due to the occlusion from facial hair.

The above variations in the dataset are probably a contributing factor as to why results for the multi-pie database are not as high as results reported for other databases (JAFFE and Cohn-Kanade). However, given the complexity of the dataset, the results are surprisingly good. Gross et al. evaluated facial expression recognition for frontal view on the multi-pie database, with results of under 50% [5]. A direct comparison with work presented in this section is unfair as the number of subjects used for training and testing was small. It should also be noted that these experiments were



Fig. 24. Examples of occlusion from facial hair present in the multi-pie database for different head poses. For columns for angles 0° and 75°, it can be seen how hair occludes the eyebrows and eyes of some subjects. However, for 90° pose the hair can cover all the features of the face.



carried out to evaluate the effect of illumination on expression recognition and not to find peak expression recognition performance. Even so, our results as high as 89% on this challenging dataset are impressive.

The results in Section 6.3.2 show the tolerance of this approach to noise introduced by the face detector (localization errors), but also to different variation in head pitch and roll angles. Psychology studies [16] have shown that head tilt (pitch angle of the head) is associated with certain groups of expressions. Pitch angles is more prominent at large yaw angles and thus may help classification of facial expressions at these angles. This could in part explain the strong performance of particular expressions at profile view over other poses.

## 7. Conclusions

The effects of pose on facial expression recognition is a largely unexplored area. Robust facial expression recognition systems must have the ability to classify expressions at different poses. This paper presents an investigation into head pose and multi-view facial expression recognition. Experiments were carried out on two databases to investigate how pose effects facial expression recognition. This paper investigates the effects of pose on facial expression recognition using variations of *LBP*s at different resolutions and different grid sampling sizes. Results in this paper have shown that *LGBP*s outperform other features. *LGBP*s utilize multi-resolution spatial histograms combined with local intensity distributions and spatial information. *LGBP*s performs particularly well at large yaw angles compared with other features. For the BU3DFE database an overall recognition rate of 67.96% is achieved for six expression over five yaw angles for four different intensity levels. *LGBP* also achieved an overall recognition rate of 80.60% for six expressions over seven yaw angles on the multi-pie dataset. *LBP<sup>ms</sup>* performed well on both databases with recognition rates of 65.02% and 73.26% respectively. *LBP<sup>ms</sup>* also showed good performance compared with more basic features like *LBP<sup>u2</sup>* and others. By comparing different features, it allows for the evaluation of the influence of orientation analysis and multi resolution analysis on facial expression recognition. Results show *LBP<sup>ms</sup>* outperforms standard *LBP<sup>u2</sup>* by up to 8% utilizing the multi-resolution analysis. Also feature *LBP<sup>u2</sup>* outperforms feature *LBP<sup>i</sup>* by about 10%, showing the importance of multi orientation analysis for facial expression recognition. Results show the strong performance of *LBP<sup>ms</sup>* and when combined with *LGBP*, a recognition rate of 71.1% is achieved on the BU3DFE database. Also this paper investigated how individual expressions performed over a range of poses. It was found that some expressions performed better at non frontal views. Also results show for some facial expressions the optimal view is feature dependent.

Experiments on the multi-pie dataset in which occlusion of some of the facial features of subjects occurs presenting a challenging dataset. Different grid sizes were investigated for extracting the feature histograms and it was shown that the local histograms outperforms a global histogram.

Results have shown the tolerance of our approach (with features *LGBP* and *LBP<sup>ms</sup>*) to misalignment errors as noise is introduced by the face detectors. In general from the above results the facial expression with most deformation have the highest recognition rates (*surprise*, *joy* and *scream*). The more subtle expressions like *squint* and *disgust* are more difficult to classify.

Experiments carried out on the BU3DFE (Section 5) have suggested that frontal view was optimal for recognition. Further still, other studies have suggested that 45° is the optimal view for facial expression recognition [7]. In summary and observing results in Table 12 and Fig. 15, experiments on both synthetic and real data

suggest that facial expression recognition is largely consistent across all poses, but the optimal view is subject to the data and features used. This is also highlighted in Tables 16 and 17 where the optimal angle can also be dependent on the grid size used to extract feature vectors.

## Acknowledgment

This work is supported by the EPSRC project LILiR (EP/E027946) and the EU project Dicta-Sign (FP7/2007–2013) under Grant No. 231135.

## References

- [1] J.C. McCall, M.M. Trivedi, Pose invariant affect analysis using thin-plate splines, in: ICPR '04: Proceedings of the Pattern Recognition, IEEE Computer Society, Washington, DC, USA, 2004, pp. 958–964.
- [2] C. Chan, J. Kittler, K. Messer, Multi-scale local binary pattern histograms for face recognition, in: The 2nd International Conference on Biometrics, ICIP 2010, pp. 809–818.
- [3] P. Ekman, W.V. Friesen, Pictures of Facial Affect, Consulting Psychologists Press, 1976.
- [4] S. Gong, P.W. McOwan, C. Shan, Dynamic facial expression recognition using a bayesian temporal manifold model, in: BMVC, vol. 1, 2006, pp. 297–306.
- [5] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-pie, Image and Vision Computing 28 (2010) 807–813 (Best of Automatic Face and Gesture Recognition 2008).
- [6] Y. Hu, Z. Zeng, L. Yin, X. Wei, J. Tu, T. Huang, Multi-view facial expression recognition, FG2008, in: 8th International Conference on Automatic Face and Gesture Recognition 2008, ICPR 2008, 2008.
- [7] Y. Hu, Z. Zeng, L. Yin, X. Wei, J. Tu, T. Huang, A study of non-frontal-view facial expressions recognition, in: 19th International Conference on Pattern Recognition 2008, ICPR 2008, 2008, pp. 1–4.
- [8] T. Kanade, Y. Tian, J.F. Cohn, Comprehensive database for facial expression analysis, in: FG '00: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000, Washington, DC, USA, IEEE Computer Society, 2000, p. 46.
- [9] S. Liao, W. Fan, A.C.S. Chung, D.Y. Yeung, Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features, in: ICIP, 2006, pp. 665–668.
- [10] S. Liao, X. Zhu, Z. Lei, L. Zhang, S. Li, Learning multi-scale block local binary patterns for face recognition, in: International Conference on Biometrics, ICBO7, 2007, pp. 828–837.
- [11] G. Littlewort, M.S. Bartlett, I. Fasel, J. Susskind, J. Movellan, Dynamics of facial expression extracted automatically from video, Journal of Image and Vision Computing, 2004, pp. 615–625.
- [12] X. Liu, J. Rittscher, T. Chen, Optimal pose for face recognition, in: CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Washington, DC, USA, 2006, pp. 1439–1446.
- [13] M.J. Lyons, J. Budynek, S. Akamatsu, Automatic classification of single facial images, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (1999) 1357–1362.
- [14] M.J. Lyons, R. Campbell, A. Plante, M. Coleman, M. Kamachi, S. Akamatsu, The non mask effect vertical viewpoint dependence of facial expression perception, Proceedings of the Royal Society, Biological Sciences 267 (2000) 2239–2245.
- [15] A. Mehrabian, Silent Messages, Wadsworth Publishing Company, Inc., Belmont, CA, 1971.
- [16] A. Mignault, A. Chaudhuri, The many faces of a neutral face: head tilt and perception of dominance and emotion, Journal of Nonverbal Behavior 27 (2003) 111–132.
- [17] S. Minut, S. Mahadevan, J.M. Henderson, F.C. Dyer, Face recognition using foveal vision, in: IEEE International Workshop on Biologically Motivated Computer Vision, 2000, pp. 424–433.
- [18] M.E.R. Nicholls, B.J. Wolfgang, D. Clode, A.K. Lindell, The effect of left and right poses on the expression of facial emotion, Neuropsychologia 40 (2002) 1662–1665.
- [19] T. Ojala, M. Pietikainen, D. Harwood, A comparative study of texture measures with classification based on feature distributions, Pattern Recognition 29 (1996) 51–59.
- [20] T. Ojala, M. Pietikainen, T. Maenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 971–987.
- [21] OpenCV, <<http://sourceforge.net/projects/opencvlibrary/>>.
- [22] 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, Part B 36 (2006) 433–449.
- [23] Y. Sato, E. Maeda, J. Yamato, K. Otsuka, S. Kumano, Pose-invariant facial expression recognition using variable-intensity templates, in: Asian Conference on Computer Vision 2007, ACCV07, pp. 1: 324–334.
- [24] C. Shan, T. Gritti, Learning discriminative lbp-histogram bins for facial expression recognition, in: Proceedings of the British Machine Vision Conference 2008.

- [25] Y. li Tian, L. Brown, A. Hampapur, S. Pankanti, A. Senior, R. Bolle, Real world real-time automatic recognition of facial expressions, in: In Proceedings of IEEE workshop on Performance Evaluation of Tracking and Surveillance, 2003.
- [26] P. Viola, M. Jones, Robust real-time object detection, *International Journal of Computer Vision* 57 (2004).
- [27] J. Wang, L. Yin, X. Wei, Y. Sun, 3D facial expression recognition based on primitive surface feature distribution, in: *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, IEEE Computer Society, 2006, pp. 1399–1406.
- [28] X. Wang, T. Han, S. Yan, An hog-lbp human detector with partial occlusion handling, in: *International Conference on Computer Vision 2009, ICCV09*, pp. 32–39.
- [29] L. Yin, X. Wei, Y. Sun, J. Wang, M. Rosato, A 3d facial expression database for facial behavior research, in: *7th International Conference on Automatic Face and Gesture Recognition 2006, FGR 2006*, 2006, pp. 211–216.
- [30] Y. Tian, T. Kanade, J. Cohn, Facial expression analysis, in: *Handbook of Face Recognition*, Springer, 2005 (Chapter 11).
- [31] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local gabor binary pattern histogram sequence (lgbphs): a non-statistical model for face representation and recognition, in: *Tenth IEEE International Conference on Computer Vision, ICCV 1*, vol. 1, 2005, pp. 786–791.
- [32] W. Zheng, H. Tang, Z. Lin, T. Huang, A novel approach to expression recognition from non-frontal face images, *Twelfth IEEE International Conference on Computer Vision, ICCV*, 2009.