

Binary Pattern Analysis for 3D Facial Action Unit Detection

Georgia Sandbach¹
gls09@imperial.ac.uk

Stefanos Zafeiriou¹
s.zafeiriou@imperial.ac.uk

Maja Pantic^{1,2}
m.pantic@imperial.ac.uk

¹ Department of Computing
Imperial College London
London, UK

² EEMCS
University of Twente
Enschede, Netherlands

Abstract

In this paper we propose new binary pattern-based features for use in the problem of 3D facial action unit (AU) detection. Two representations of 3D facial geometries are employed, the depth map and the Azimuthal Projection Distance Image (APDI). To these the traditional Local Binary Pattern is applied, along with Local Phase Quantisation, Gabor filters and Monogenic filters, followed by the binary pattern feature extraction method. Feature vectors are formed for each feature type through concatenation of histograms formed from the resulting binary numbers. Feature selection is then performed using a two-stage GentleBoost approach. Finally, we apply Support Vector Machines as classifiers for detection of each AU. This system is tested in two ways. First we perform 10-fold cross-validation on the Bosphorus database, and then we perform the cross-database testing by training on this database and then testing on apex frames from the D3DFACS database, achieving promising results in both.

1 Introduction

Recognition of facial expressions is a challenging problem, as the face is capable of complex motions, and the range of possible expressions is extremely wide. For this reason, detection of facial action units (AUs) from the Facial Action Coding System has become a widely studied area of research. AUs are the building blocks of expressions, and are finite in number, thus allowing a comprehensive detection system to be produced.

3D facial geometry data is an relatively new area of expression recognition research. Data of this kind allow a greater amount of information to be captured, including out-of-plane movement which 2D cannot record, whilst also removing the problems of illumination and pose inherent to 2D data. 3D data has previously been employed for full facial expression recognition (e.g. [17]) and also facial AU detection (e.g. [20]). The majority of research conducted on static 3D facial meshes so far has employed features based on facial points (e.g. [9, 21, 24]), patches on the mesh (e.g. [8, 11]) or morphable models (e.g. [12, 32]). Alternatively, 2D representations such as the depth map and curvature images have been exploited (e.g. [2, 20]). Further details of the methods employed can be found in [18]. This work aims to explore alternative feature types suitable for analysis of 3D facial expressions

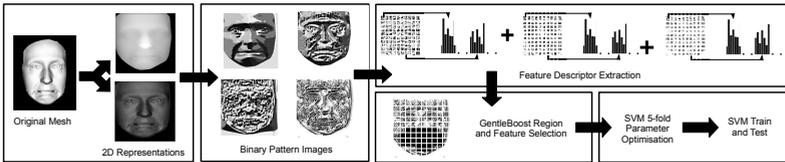


Figure 1: An overview of our proposed system.

by exploiting the popular binary pattern approach which has been successfully applied to the 2D problem.

Local Binary Patterns (LBPs) [13] are a technique that have been widely applied to the problems of facial expression recognition [22] and face recognition [1]. They are a useful feature type as they simply and quickly encode the shape of the image at each pixel by computing binary numbers that reflect the local neighbourhood. Variants on LBPs have also been proposed to improve on the performance of the basic feature: Local Gradient Orientation Binary Patterns (LGOBPs) [10], Local Phase Quantisers (LPQ) [14], Local Gabor Binary Patterns (LGBPs) [30], and Histogram of Monogenic Binary Patterns (HMBPs) [29]. LBPs, LPQs and LGBPs have also been extended to the dynamic problem in the form of LBP-TOP [31], LPQ-TOP [7] and V-LGBPs [28]. The traditional LBP descriptor has also been applied to the depth map of a 3D facial mesh in 3DLBPs [6] and the Multi-resolution Extended Local Binary Pattern (MELBPs) [5]. More recently we have proposed Local Normal Binary Patterns (LNBP) [16] to employ the facial mesh normals for 3D AU detection.

In this paper, we explore further ways of exploiting the 3D facial geometry information through the use of binary pattern methods. In order to apply these algorithms, we transform the 3D facial geometry into two 2D representations that contain the geometry information. Firstly, we employ the depth map representation with a variety of feature types, phase quantisers, Gabor and Monogenic filters, in order to produce new feature types. Secondly, we utilise a new 2D representation, the Azimuthal Projection Distance Image, which represents the direction of the normals in the facial geometry thus capturing different geometry information. In summary we propose seven new feature types: (1) Local Azimuthal Binary Patterns (LABPs) (2) Local Depth Phase Quantisers (LDPQs) (3) Local Azimuthal Phase Quantisers (LAPQs) (4) Local Depth Gabor Binary Patterns (LDGBPs) (5) Local Azimuthal Gabor Binary Patterns (LAGBPs) (6) Local Depth Monogenic Binary Patterns (LDMBPs) (7) Local Azimuthal Monogenic Binary Patterns (LAMBPs). Each of these is employed with two-stage GentleBoost feature selection and Support Vector Machine (SVM) classification in order to test their effectiveness as compared to the original 3DLBP feature and our LNBP feature type, including conducting the first cross-database testing carried out on 3D AUs. An overview of our system can be seen in Fig. 1.

2 Facial Geometry Representations

We examine the use of two different 2D representations of the 3D facial geometry: the widely used depth map representation, and the Azimuthal Projection Distance Image (APDI) which measures the Euclidean distance of the facial mesh normals at each point.

The first representation, the depth map, is widely used in 3D facial analysis [2, 5, 26] as it is a very simple 2D representation. In this work, a regular grid is defined with suitable x

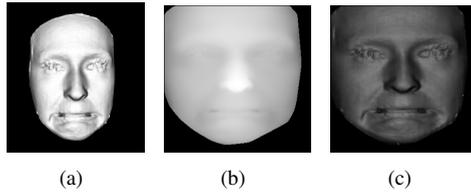


Figure 2: 2D representations of the facial mesh for subject bs043 from the Bosphorus database performing AU20 (a) The original facial mesh (b) The interpolated depth map (b) The Azimuthal Projection Distance Image.

and y ranges, and Delaunay triangulation is used to interpolate the height of the facial mesh points in the x - y plane within this range so that a regular grid of z values are generated which forms the depth map. An example of this can be seen in Fig. 2(b) for the facial mesh in Fig. 2(a).

As alternative, we outline a second representation, the Azimuthal Projection Distance Image (APDI). The aim of this method is to find a representation that allows accurate comparison of the directions of the normals in the local neighbourhood. The Azimuthal Equidistant Projection (AEP) is able to project normals onto points in a Euclidean space according to the direction. It has previously been employed to capture the local variations in facial shape [23], and can be applied to the normals, in order to project each 3D direction into the position in a Euclidean 2D plane. So for a regular grid of normals, defined as $\mathbf{n}(i, j) = (u_{i,j}, v_{i,j}, w_{i,j})$, the AEP point $\mathbf{p}(i, j) = (x_{i,j}, y_{i,j})$ in this plane is defined as:

$$\begin{aligned} x_{i,j} &= k' \cos\theta(i, j) \sin[\phi(i, j) - \hat{\phi}(i, j)] \\ y_{i,j} &= k' (\cos\hat{\theta}(i, j) \sin\phi(i, j) - \sin\hat{\theta}(i, j) \cos\theta(i, j) \cos[\phi(i, j) - \hat{\phi}(i, j)]) \end{aligned} \quad (1)$$

where $\theta(i, j) = \frac{\pi}{2} - \arcsin(w_{i,j})$ is the elevation angle measured from the z -axis, $\phi(i, j) = \arctan \frac{v_{i,j}}{u_{i,j}}$ is the azimuth angle, $\hat{\theta}(i, j)$ and $\hat{\phi}(i, j)$ are the elevation and azimuth of the mean normal $\hat{\mathbf{n}}(i, j)$ at the point \mathbf{p} , $k' = \frac{c}{\sin(c)}$ where c is defined such that

$$\cos(c) = \sin\hat{\theta}(i, j) \sin\theta(i, j) + \cos\hat{\theta}(i, j) \cos\theta(i, j) \cos[\phi(i, j) - \hat{\phi}(i, j)] \quad (2)$$

However, for our purposes, it is necessary to be able to directly compare the projection coordinates of neighbouring points, and so for this reason $\hat{\theta}$ and $\hat{\phi}$ at every point are set to be $\frac{\pi}{2}$ and 0 respectively, so that the distance calculated is always compared to a normal $\hat{\mathbf{n}} = (1, 0, 0)$ which was chosen as a reference to create an image suitable for further analysis. This assumption makes $\cos(c) = \sin\theta(i, j)$ and allows the projection to be simplified to:

$$x_{i,j} = k' \cos\theta(i, j) \sin\phi(i, j) \quad y_{i,j} = k' \cos\theta(i, j) \cos\phi(i, j) \quad (3)$$

The above formulation then allows distances between the normals in Euclidean space to be directly found, and this simplification also reduces the complexity of the feature extraction process. In order to employ this in the binary pattern framework, the coordinates are used to find an absolute distance from the origin $d_{i,j} = \sqrt{x_{i,j}^2 + y_{i,j}^2}$, and these values form the Azimuthal Projection Distance Image (APDI) for the facial mesh. An example of this calculated for the facial mesh seen in Fig. 2(a) can be seen in Fig. 2(c).

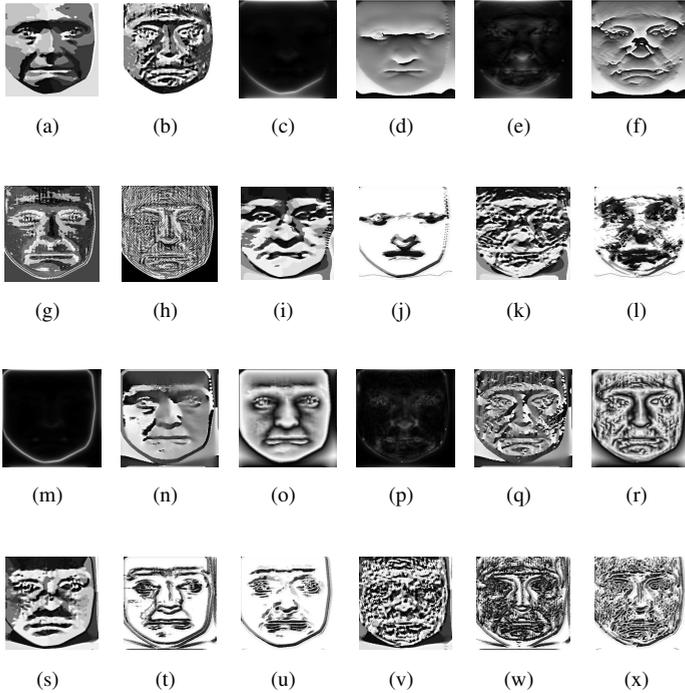


Figure 3: Examples of images produced by feature descriptors for subject bs043 performing AU20. M - magnitude, P - phase, O - orientation. (a) 3DLBP image (b) LABP image (c)-(d) Gabor depth M/P images (e)-(f) Gabor APDI images (g) LDPQ image (h) LAPQ image (i)-(j) LDGBP M/P images (k)-(l) LAGBP images (m)-(o) Monogenic depth M/P/O images (p)-(r) Monogenic APDI images (s)-(u) LDMBP M/P/O images (v)-(x) LMBBP images.

3 Binary Pattern Features

In this section we explain the proposed set of new binary pattern features for analysis of 3D facial geometry information. First, the original 3DLBP feature is described, and its extension using the APDI. Then we describe the application of Local Phase Quantisers, and Gabor and Monogenic binary patterns, to 3D facial meshes, both utilising the depth map and APDI.

3.1 Local Binary Patterns

3D Local Binary Patterns (3DLBPs) were proposed for use in facial expression recognition of 3D meshes in [6]. They exploit the depth map representation of the 3D information, interpolated onto a regular grid, in order to encode the local shape around each point in the mesh. In general, for an image I , a neighbourhood is defined as a circle around each pixel with a radius r and number of points P spaced at regular angles around the circle. The central pixel value is then used as a threshold to assign binary bits to the pixels in the neighbourhood, thus producing a binary number for that pixel:

$$LBP(x_c, y_c) = \sum_{p=0}^P 2^p s(I(x_p, y_p) - I(x_c, y_c)) \quad \text{where} \quad s(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Here we propose a new feature based on this idea, the Local Azimuthal Binary Pattern (LABP). This feature employs the APDI rather than the depth map to calculate the binary pattern in the neighbourhood of each point. In this way, the shape of the mesh is encoded via the direction of the normals at each point, allowing more subtle information to be captured about the structure of the mesh. This method is similar to gradient-based methods employed in 2D image processing [10]. Examples of the 3DLBP and LABP images resulting from the examples in Figs. 2(b) and 2(c) can be seen in Figs. 3(a) and 3(b) respectively.

3.2 Local Phase Quantisers

Local Phase Quantisers (LPQs) can be used to extract the local phase information, which contains important directional features (e.g. edges) [15] useful in analysis of the facial deformations, from the image using the Short-Term Fourier Transform (STFT) calculated over a rectangular neighbourhood, and are designed to be particularly invariant to blurring as this phase information is unaffected by an assumed symmetric blur pattern [14].

The standard LPQ feature is extracted as follows. For a point $\mathbf{p} = (x, y)$ with $M \times M$ neighbourhood $\mathcal{N}_{\mathbf{p}}$, the STFT is defined as:

$$F(\mathbf{u}, \mathbf{p}) = \sum_{\mathbf{k} \in \mathcal{N}_{\mathbf{p}}} f(\mathbf{p} - \mathbf{k}) e^{-j2\pi \mathbf{p}^T \mathbf{k}} = \mathbf{w}_{\mathbf{u}}^T \mathbf{f}_{\mathbf{p}} \quad (5)$$

where $\mathbf{w}_{\mathbf{u}}$ is the basis vector of the 2D DFT at frequency \mathbf{u} and $\mathbf{f}_{\mathbf{p}}$ is the vector form of the neighbourhood $\mathcal{N}_{\mathbf{p}}$. In addition, in our implementation a Gaussian window is also applied to the basis functions. Four frequency pairs are employed, $([0, 0]^T, [0, a]^T)$, $([a, a]^T, [a, -a]^T)$, and these allow the calculation of two binary digits each in order to form the eight-digit binary number for the point \mathbf{p} in the following way:

$$q(\mathbf{p}) = [s(\text{Re}\{F(\mathbf{u}_1, \mathbf{p})\}), \dots, s(\text{Re}\{F(\mathbf{u}_4, \mathbf{p})\}), s(\text{Im}\{F(\mathbf{u}_1, \mathbf{p})\}), \dots, s(\text{Im}\{F(\mathbf{u}_4, \mathbf{p})\})] \quad (6)$$

where $s(v)$ is as defined in the previous section.

Here we extend this idea by application of the LPQ feature to our two representations, the depth map and APDI, in order to produce two operators, Local Depth Phase Quantisers (LDPQs) and Local Azimuthal Phase Quantisers (LAPQs). Examples of the resulting LDPQ and LAPQ images can be seen in Figs. 3(g) and 3(h) respectively.

3.3 Gabor Binary Patterns

Gabor filters have been widely used 2D facial expression recognition [25, 27], and also applied to 3D analysis [20], as they are well suited to capturing the structural information, namely edge features, in an image in a way that is similar to the human visual system. In this work we employ log Gabor filters of various scales and orientations.

The transfer function of the filters used consists of a radial log Gabor filter multiplied by an angular Gaussian component:

$$G(\mathbf{u})_{\nu, \theta} = \exp\left(\frac{-(\log(\nu|\mathbf{u}|))^2}{2(\log \sigma)^2}\right) \exp\left(\frac{-(\angle \mathbf{u} - \theta)^2}{2\sigma_{\phi}}\right) \quad (7)$$

where ν and θ are the scale and orientation of the filter respectively, and σ and σ_{ϕ} define the spread of the filter in the radial and angular directions respectively.

We multiply this to the Fourier transform of our two images, the depth map and APDI, at four different scales and four orientations, and then the inverse transform is taken to find

the resulting Gabor coefficients, $g(x, y) = F^{-1}(GF(I))$. The magnitude and phase of these are then taken as new images, $g_M(x, y) = |g(x, y)|$ and $g_P(x, y) = \angle g(x, y)$. The binary pattern algorithm is then applied to each of the resulting magnitude images in order to encode the local structural information further, as was done in the 2D LGBP case [30]. This forms the magnitude half of each of our new features, Local Depth Gabor Binary Patterns (LDGBPs) and Local Azimuthal Binary Patterns (LAGBPs).

In addition, here we also encode the phase information in each image using a variant on the LBP method due to the circular nature of phase. In this case the difference is taken between the phase at the central point and those of the neighbouring points and this difference is compared to a threshold value, ψ , in order to assign a zero or one for each neighbouring point. For these experiments this threshold value was set to be $\frac{\pi}{4}$. For example, the $LDGBP_P$ operator, applied to the phase of the depth Gabor features, is calculated as:

$$LDGBP_P(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(\psi - |g_P(x_c, y_c) - g_P(x_p, y_p)|) \quad (8)$$

where $g_P(x_c, y_c)$ is the phase image at the central point, $g_P(x_p, y_p)$ is the phase image at the p^{th} point in the neighbourhood and P is the number of points in the neighbourhood. The $LAGBP_P$ operator is similarly formed, but applied to the phase of the APDI Gabor features. Examples of the Gabor images for the depth map and APDI, and the corresponding LDGBP and LAGBP images, can be seen in Figs. 3(c)-3(d), 3(e)-3(f), 3(i)-3(j), and 3(k)-3(l) respectively.

3.4 Monogenic Binary Patterns

The monogenic signal is an alternative approach to the use of Gabor filters. It allows a 2D image to be analysed in terms of magnitude, phase and orientation, thus giving a 2D representation of the structural information in the data. Therefore, it is no longer necessary to apply filters at multiple orientations, as with Gabor filters, as features in multiple directions are captured simultaneously. However, multiple scales are still useful for capturing different levels of structure in the image. The monogenic representation is a 2D version of the analytic signal, which exploits the Riesz transform in order to achieve the same properties as the 1D version [4].

In practise, the three components of the monogenic signal, magnitude, m_M , phase, m_P , and orientation, m_O , can be calculated through the use of two orthogonal monogenic filters with transfer functions $H_1(\mathbf{u}) = \frac{j|u_1|}{|\mathbf{u}|}$ and $H_2(\mathbf{u}) = \frac{j|u_2|}{|\mathbf{u}|}$, and radial log Gabor filters with varying scales:

$$m_M(x, y) = \sqrt{g'(x, y)^2 + h'_1(x, y)^2 + h'_2(x, y)^2} \quad (9)$$

$$m_P(x, y) = \arctan\left(\frac{h'_2(x, y)}{h'_1(x, y)}\right) \quad m_O(x, y) = \arctan\left(\frac{g'(x, y)}{\sqrt{h'_1^2 + h'_2^2}}\right)$$

where $h'_i = F^{-1}(H_i G' F(I))$, I is the image, F is the 2D Fourier transform, and $G'(\mathbf{u})_v = \exp(-(\log(v|\mathbf{u}|))^2 / 2(\log \sigma)^2)$ is the transfer function of the radial component of the full log Gabor filter outlined in the previous section and $g' = F^{-1}(G' F(I))$.

This signal is calculated for both the depth map and APDI, in order to produce the Local Depth Monogenic Binary Pattern (LDMBP) and Local Azimuthal Monogenic Binary Pattern (LAMBP) respectively. The magnitude, phase and orientation images are then encoded using

the binary pattern algorithm and phase variation described in equation 8. This contrasts with the original 2D implementation of the HMBP [29], as in that work only the magnitude and phase were encoded, so that the second dimension of structural information was lost. In addition, the phase variant of the LBP employed here encodes the shape purely based on differences between phase, not similar quadrants. Examples of the Monogenic images for the depth map and APDI, and the corresponding LDMBP and LAMB images, can be seen in Figs. 3(m)-3(o), 3(p)-3(r), 3(s)-3(u), and 3(v)-3(x) respectively.

4 Methodology

In order to extract features using the previously outlined feature types, the 3D facial meshes are first aligned and used to create the depth maps and APDIs. Alignment is achieved via six landmarks in the face, given in the Bosphorus database and manually selected in the D3DFACS database, by applying a calculated rotation, translation and stretch in the x and y directions to ensure correspondence. Once this is done the depth map and APDI can be found from interpolated grids of the depth and normals, and feature extraction is then performed using each of the feature types with 8 set as the value of both the radius and number of neighbourhood points for each of the operators. Feature vectors are created for each of the above descriptors through the use of histograms. First, the x - y plane of the mesh is divided into 10×10 equally-sized square blocks, and for each of these a histogram is formed from the calculated binary numbers. These histograms are then concatenated into one large feature vector. 60 bins were used to produce the feature descriptors. In the case of LDGBPs, LAGBPs, LDMBPs and LAMBPs, multiple images will be used, for magnitude and phase and each scale and orientation, and so will generate a separate histogram. These are formed into one feature descriptor via further histogram concatenation. This process can be seen in Fig. 4.

Feature selection is performed in order to reduce the dimensionality of the feature vectors before classification. GentleBoost, a modified version of the AdaBoost algorithm which gives a more stable result than the original, was used for this purpose, with two stages to the feature selection. The first stage consisted of choosing regions in the image which gave the most discriminative information about the examples for the AU. At each stage the error rate of the features within the regions when classified by a weak classifier are averaged, and region with the lowest chosen. Then all features within the region are used to update the examples weightings at that stage. This step allowed features in parts of the face where the AU is not active to be discarded quickly before the main feature selection step. The second stage is then a feature-wise GentleBoost selection algorithm to choose particular features within those regions. To avoid overfitting, our strategy is to run this stage of the selection algorithm repeatedly, removing the previously chosen features at each stage, until the number

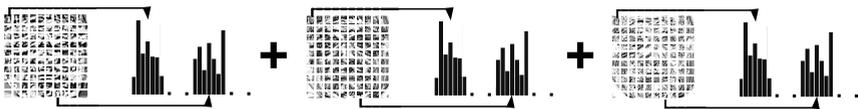


Figure 4: Construction of the feature vector by concatenation of histograms from each block in the image, and then concatenation of different image histograms, in this case the magnitude, phase and orientation images for once scale of the LAMBPs.

of features selected exceeds the number of examples in the training set, or until fewer than 5 features are being chosen by the algorithm.

SVM classifiers were then trained for detection of each AU. These classifiers employed the histogram intersection kernel, and parameter optimisation was first performed using 5-fold cross-validation on a separate validation set which was taken to be one third of the data available for training. The SVM classifiers were then trained on the remaining training set.

5 Experimental Results

Experiments were conducted to compare the performance of the different feature types on two 3D AU detection problems to both the 3DLBP feature type and also to our previously proposed Local Normal Binary Patterns (LNBP) feature type [16]. Here we only use the $LNBP_{OA}$ operator for testing as this generally outperformed the $LNBP_{TA}$ operator in our previous work. The first test consisted of 10-fold cross-validation performed on the Bosphorus database [19], that consists of static images of 105 subjects performing up to 24 of the AUs. Secondly, the D3DFACS database [3] was employed in order to perform cross-database testing, which is the first time tests of these kind have been performed on 3D AUs. This database consists of 10 subjects performing a wide range of single and combinations of AUs. The AUs that are present in both databases were identified and classifiers were trained on the Bosphorus database for each of these AUs, and then testing was conducted on an apex frame from each sequence in the D3DFACS database that contained one or more of the AUs. This latter test poses a difficult problem for the system, as the D3DFACS contains examples of many combinations of AUs, whereas the Bosphorus database only contains examples of single AUs being performed.

5.1 Cross-Validation Testing

The ROC Area under the Curve (AuC) is used to evaluate the performance of the system with the different feature types. Results from Bosphorus database cross-validation testing can be seen on left side of Table 1. The left column shows the results for 3DLBPs, while the other seven columns display our new proposed feature types. As can be seen, on average LDPQs, LDGBPs, LAGBPs, LDMBPs, and LAMBPs all out perform 3DLBPs, while LAPQs and LABPs achieve the same performance on average. However, the benefits of the latter feature type can be seen by looking at individual AU results such as AUs 1, 12, and 17, plus 15 and 20 where LABPs significantly outperform 3DLBPs and do better than LNBP which suggests that the APDI image is a better method for comparison of normals than the LNBP approach. LAPQs only achieve a benefit for a couple of AUs, 17 and 20, achieving a significant improvement in the latter. The highest result was achieved when employing the LDGBP feature type, with AuC of 97.2. This compares favourably with results achieved on this same database in [20] in which a maximum AuC of 96.3 was attained when using multiple 2D representations of 3D facial geometries with Gabor features. Several other feature types also performed very well, LDPQs, LAGBPs, LDMBPs and LAMBPs. They all achieved a significant increase on average over 3DLBPs and in one or more AUs. In addition, the azimuthal features outperform their depth counterparts in a number of cases: LAGBPs for AUs 16, 20 and 34, and LAMBPs in 15, 20. However, generally the depth and azimuthal features match each other for performance. This suggests that some of the benefits of the APDI seen in the LABP results are lost in the more complex features.

AU	Cross-Validation Results									Cross-Database Results								
	B	NB	AB	DPQ	APQ	DGB	AGB	DMB	AMB	B	NB	AB	DPQ	APQ	DGB	AGB	DMB	AMB
1	92.2	96.4	95.1	95.6	93.4	95.7	94.2	94.0	93.5	86.0	91.8	81.3	88.3	83.3	85.4	86.2	90.1	88.0
2	98.0	97.9	98.5	98.5	98.7	99.0	99.0	98.7	98.9	94.2	97.2	94.6	95.6	92.9	98.8	87.2	99.2	91.6
4	96.3	96.1	96.9	96.4	96.3	97.9	97.6	97.5	97.5	85.9	88.6	80.4	86.4	79.4	90.0	80.5	86.3	89.5
43	99.6	99.2	99.8	99.8	99.7	99.9	100.0	99.9	100.0	99.0	97.3	95.7	98.6	91.2	99.1	99.4	98.5	97.4
44	93.9	88.8	91.4	95.7	92.4	95.7	95.3	94.6	93.7	97.8	97.8	97.0	96.0	99.1	98.3	97.2	96.5	92.9
9	97.8	96.8	97.7	97.6	97.6	98.6	97.6	98.1	97.3	93.9	95.8	91.6	95.5	87.4	96.1	82.7	95.5	85.4
10	97.8	96.9	96.1	96.5	96.5	97.6	97.3	97.2	97.3	83.1	90.0	88.8	90.2	88.2	88.9	85.8	89.1	88.7
12	94.9	96.4	96.5	96.4	96.0	96.7	96.5	97.1	95.6	90.4	93.9	92.6	91.5	88.2	94.3	94.4	94.8	94.2
12L	96.9	98.9	95.8	97.9	95.9	98.2	97.6	98.2	97.1	N	N	N	N	N	N	N	N	N
12R	97.6	99.4	97.4	97.6	97.1	98.3	97.0	97.9	96.1	N	N	N	N	N	N	N	N	N
14	92.4	92.9	92.1	94.6	92.4	95.7	93.9	95.9	94.6	91.4	90.5	72.7	91.1	85.5	92.2	89.3	89.3	89.2
15	85.8	90.1	91.7	90.3	86.6	92.7	90.1	89.9	92.4	72.1	75.7	86.0	91.9	81.6	85.9	71.4	89.8	79.5
16	96.4	93.5	94.4	95.2	95.4	96.7	98.1	96.9	94.2	92.2	82.5	77.2	94.1	71.9	90.9	47.6	92.3	68.7
17	93.4	95.5	94.6	96.6	95.1	96.9	96.1	96.7	95.5	84.1	83.5	64.7	87.9	65.2	83.8	70.4	85.2	68.0
18	97.0	96.2	97.3	97.5	95.5	98.2	97.8	98.1	97.7	93.4	91.7	86.2	98.3	86.8	96.9	84.6	98.5	83.3
20	89.9	92.9	96.3	94.9	94.8	95.1	96.3	94.1	95.7	83.0	88.6	83.1	90.7	89.0	88.7	74.0	85.0	73.7
22	99.3	98.0	98.5	99.2	98.0	99.6	99.1	99.7	98.6	95.1	95.5	89.9	96.2	82.2	97.4	90.1	97.6	89.8
23	94.6	90.8	94.2	96.4	95.2	96.4	94.3	95.9	93.5	51.9	48.9	46.6	49.6	49.2	50.1	45.4	49.1	47.7
24	88.8	89.1	86.2	92.0	87.6	92.8	91.3	93.8	90.3	76.0	72.8	71.4	79.8	68.4	85.8	72.4	74.9	71.2
25	94.8	92.5	92.6	94.0	89.7	95.4	92.3	95.8	95.2	74.7	70.8	61.9	68.3	38.0	78.0	47.0	79.5	38.3
26	93.8	93.4	94.5	95.7	94.0	96.6	97.0	97.0	95.0	71.3	76.4	63.4	78.5	67.5	75.0	74.4	76.4	72.3
27	99.5	97.9	99.1	99.5	99.3	99.7	98.6	99.8	99.1	94.6	93.0	92.9	96.3	87.3	94.4	90.5	92.7	92.4
28	97.8	97.7	97.9	98.8	98.2	99.1	99.0	99.3	99.2	98.9	99.9	99.4	99.5	91.4	99.9	98.8	100.0	99.6
34	99.0	97.0	98.4	99.7	98.9	99.3	99.7	99.1	99.0	97.3	95.2	99.7	97.2	95.3	98.1	99.1	98.1	76.7
μ	95.3	95.2	95.5	96.5	95.2	97.2	96.5	96.9	96.1	86.6	87.1	82.6	88.9	80.4	89.4	80.3	89.0	80.8
σ	0.7	0.8	0.6	0.5	0.7	0.4	0.5	0.5	0.6									

Table 1: ROC AuC Results. Left: Cross-validation testing on the Bosphorus database. Right: Cross-database testing trained on the Bosphorus database and tested on apex frames from the D3DFACS database. B: 3DLBPs, NB: LNBPs, AB: LABPs, DPQ: LDPQs, APQ: LAPQs, DGB: LDGBPs, AGB: LAGBPs, DMB: LDMBPs, AMB: LAMBPs. μ : Mean AuC score across AUs, σ : Standard deviation of mean AuC score across folds. N: No examples.

5.2 Cross-Database Testing

The results from cross-database testing on the D3DFACS database can be seen on the right side of Table 1. The first thing of note about these results is that again it is LDGBPs that achieve the highest result, of 89.4. However, for the majority of AUs it does not achieve the highest result, as it is outperformed by either LDMBPs or LDPQs, which both also achieve high average scores. It obvious from these results that the azimuthal binary pattern features all do far worse than their depth counterparts when tested on this database. LABPs actually achieve the highest result of the four, and does better than 3DLBPs on a number of AUs, noticeably 10 and 15. However, it achieves the same performance or lower on a larger number of AUs, many of which it outperformed 3DLBPs in the cross-validation test such as 1 and 20. It also does worse than LNBPs in this test, suggesting that it does not generalise as well to the new database. This could be due to the fact that the LABP feature type will be much more sensitive to differences between the smoothness in the facial meshes, both across the training database, and between the training and testing databases. This problem is also reflected in the LAGBP and LAMBPs results, where the features fail to achieve similar or better results than their depth counterparts on the majority of AUs. This along with the cross-validation results suggests that these complex features are even more sensitive to the mesh smoothness differences.

6 Conclusions

In this paper we have proposed a series of new feature types suitable for AU detection in 3D facial meshes. These are based on two representations of the 3D information, the previously employed depth map and the APDI. Experimental testing was conducted first on a single database using cross-validation, and then cross-database testing using a second database.

The results show that while the majority of the features outperformed the standard 3DLBP on the single database, the APDI-based features struggled to generalise to a new database of examples. This could be because of the differences in smoothness between the two databases, and future work will need to focus on improving how robust these features are to changes of this nature.

7 Acknowledgements

The research presented in this paper has been funded by the European Research Council under the ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB). The work of G. Sandbach is funded by the Engineering and Physical Sciences Research Council through a DTA studentship. The work of S. Zafeiriou has been partially funded by an Imperial College London Junior Research Fellowship.

References

- [1] T. Ahonen, A. Hadid, et al. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 2037–2041, 2006. ISSN 0162-8828.
- [2] Stefano Berretti, Boulbaba Ben Amor, Mohamed Daoudi, and Alberto del Bimbo. 3D facial expression recognition using SIFT descriptors of automatically detected keypoints. *The Visual Computer*, 27:1021–1036, 2011. ISSN 0178-2789. URL <http://dx.doi.org/10.1007/s00371-011-0611-x>.
- [3] D. Cosker, E. Krumbhuber, and A. Hilton. A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2296–2303. IEEE, 2011.
- [4] M. Felsberg and G. Sommer. The monogenic signal. *Signal Processing, IEEE Transactions on*, 49(12):3136–3144, 2001.
- [5] D. Huang, M. Ardabilian, Y. Wang, and L. Chen. A novel geometric facial representation based on multi-scale extended local binary patterns. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 1–7. IEEE, 2011.
- [6] Y. Huang, Y. Wang, and T. Tan. Combining statistics of geometrical and correlative features for 3D face recognition. In *Proceedings of the British Machine Vision Conference*, pages 879–888. Citeseer, 2006.
- [7] 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)*, Santa Barbara, CA, USA, March 2011.
- [8] Pierre Lemaire, Boulbaba Ben Amor, Mohsen Ardabilian, Liming Chen, and Mohamed Daoudi. Fully automatic 3D facial expression recognition using a region-based approach. In *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*, J-HGBU '11, pages 53–58, New York, NY, USA,

2011. ACM. ISBN 978-1-4503-0998-1. doi: 10.1145/2072572.2072589. URL <http://doi.acm.org/10.1145/2072572.2072589>.
- [9] X. Li, Q. Ruan, and Y. Ming. 3D Facial expression recognition based on basic geometric features. In *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*, pages 1366–1369, oct. 2010.
- [10] S. Liao and A. Chung. Face recognition with salient local gradient orientation binary patterns. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 3317–3320. IEEE, 2010.
- [11] A. Maalej, B.B. Amor, M. Daoudi, A. Srivastava, and S. Berretti. Shape analysis of local facial patches for 3D facial expression recognition. *Pattern Recognition*, 44(8): 1581–1589, 2011.
- [12] S. Mpiperis, S. Malassiotis, and M. G. Strintzis. Bilinear decomposition of 3D face images: an application to facial expression recognition. In *10th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2009)*, 2009.
- [13] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, pages 971–987, 2002.
- [14] V. Ojansivu and J. Heikkilä. Blur insensitive texture classification using local phase quantization. *Image and Signal Processing*, pages 236–243, 2008.
- [15] A.V. Oppenheim and J.S. Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 1981.
- [16] G. Sandbach, S. Zafeiriou, and M. Pantic. Local Normal Binary Patterns for 3D Facial Expression Recognition. In *International Conference on Image Processing (ICIP 2012)*, October 2012. Accepted for publication.
- [17] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert. Recognition of 3D facial expression dynamics. *Image and Vision Computing*, 2012. ISSN 0262-8856. doi: 10.1016/j.imavis.2012.01.006.
- [18] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin. Static and dynamic 3D facial expression recognition: A comprehensive survey. *Image and Vision Computing*, 2012. ISSN 0262-8856. doi: 10.1016/j.imavis.2012.06.005.
- [19] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun. Bosphorus database for 3D face analysis. *Biometrics and Identity Management*, pages 47–56, 2008.
- [20] A. Savran, B. Sankur, and M. Taha Bilge. Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units. *Pattern recognition*, 45(2):767–782, 2012.
- [21] T. Sha, M. Song, J. Bu, C. Chen, and D. Tao. Feature level analysis for 3D facial expression recognition. *Neurocomputing*, 2011.

- [22] C. Shan, S. Gong, and P.W. McOwan. Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816, 2009. ISSN 0262-8856.
- [23] W.A.P. Smith and E.R. Hancock. Recovering facial shape using a statistical model of surface normal direction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):1914–1930, 2006.
- [24] H. Soyel and H. Demirel. Optimal feature selection for 3D facial expression recognition using coarse-to-fine classification. *Turkish Journal of Electrical Engineering and Computer Sciences*, 18(6):1031–1040, 2010.
- [25] Y. Tong, W. Liao, and Q. Ji. Facial action unit recognition by exploiting their dynamic and semantic relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1683–1699, 2007.
- [26] N. Vretos, N. Nikolaidis, and I. Pitas. 3d facial expression recognition using zernike moments on depth images. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 773–776, sept. 2011. doi: 10.1109/ICIP.2011.6116669.
- [27] T. Wu, M.S. Bartlett, and J.R. Movellan. Facial expression recognition using Gabor motion energy filters. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 42–47. IEEE, 2010.
- [28] S. Xie, S. Shan, X. Chen, and W. Gao. V-LGBP: Volume based local Gabor binary patterns for face representation and recognition. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2009.
- [29] M. Yang, L. Zhang, L. Zhang, and D. Zhang. Monogenic binary pattern (mbp): A novel feature extraction and representation model for face recognition. In *2010 International Conference on Pattern Recognition*, pages 2680–2683. Ieee, 2010.
- [30] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 786–791. IEEE, 2005. ISBN 076952334X.
- [31] G. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 915–928, 2007.
- [32] X. Zhao, E.D. Di Huang, and L. Chen. Automatic 3D facial expression recognition based on a Bayesian Belief Net and a Statistical Facial Feature Model. In *2010 International Conference on Pattern Recognition*, pages 3724–3727, 2010.