LOCAL NORMAL BINARY PATTERNS FOR 3D FACIAL ACTION UNIT DETECTION

Georgia Sandbach[*], *Stefanos Zafeiriou*[*] *and Maja Pantic*[*†]

[*]Imperial College London, Department of Computing
180 Queen's Gate, London, SW7 2AZ
[†]University of Twente, EEMCS, Netherlands

ABSTRACT

This paper proposes a new feature descriptor, local normal binary patterns (LNBPs), which is exploited for detection of facial action units (AUs). After LNBPs have been employed to form descriptor vectors, which capture the detailed shape of the action, feature selection is performed via a GentleBoost (GB) algorithm, and support vector machines (SVMs) are trained to detect each AU. This process was tested on the Bosphorus database, alongside the same test using 3D local binary pattern (3DLBP) descriptors which apply the LBP operator to the depth map of the face. LNBP descriptors were demonstrated to outperform 3DLBPs in detection of many individual AUs. Finally, feature fusion was used to combine the benefits of the 3DLBPs and each of the LNBP descriptors, with the best result achieving a mean ROC AuC of 96.35.

Index Terms— local normal binary patterns, facial action units, expressions, 3D facial geometry, support vector machines.

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 so allow for complex analysis of the full range of expressions, while being finite in number, thus allowing a comprehensive detection system to be produced.

The use of 3D facial geometry data and extracted 3D features for expression recognition has so far not been heavily studied. Images and videos of this kind allow a greater amount of information to be captured (2D and 3D), including out-of-plane movement which 2D cannot record, whilst also removing the problems of illumination and pose inherent to 2D data. For this reason some work has begun to employ 3D facial geometry data for facial expression recognition, for example [1]. 3D facial geometries have also started to be employed for facial AU detection, for example in [2] which uses curvature images as 2D representations of the 3D data for feature extraction.

Local Binary Patterns (LBPs) are a technique that have been widely applied to the problem of facial expression recognition [3]. Variants on LBPs have also been proposed: local gradient orientation binary patterns (LGOBPs) [4], local phase quantisers (LPQ)[5], and local Gabor binary patterns (LGBPs) [6]. LBPs, LPQs and LGBPs have also been extended to the dynamic problem in the form of LBP-TOP [7], LPQ-TOP [8] and V-LGBPs [9]. Some works have also begun to apply descriptors of these kind to the 3D problem by proposing the 3DLBP, the traditional LBP descriptor applied to the depth map of a facial mesh [10] and the Multiresolution Extended Local Binary Pattern (MELBPs) [11].

This paper proposes a new type of feature descriptor, local normal binary patterns (LNBPs), that can be applied to 3D meshes to encode the shape of the mesh at every point. We propose two novel LNBP feature descriptors for use on 3D facial geometries. These descriptors are tested alongside the 3DLBP feature descriptor, which is the traditional LBP applied to the depth map of the 3D facial geometries as comparison. In addition the benefits of the different descriptors are harnessed by combining them through feature fusion.

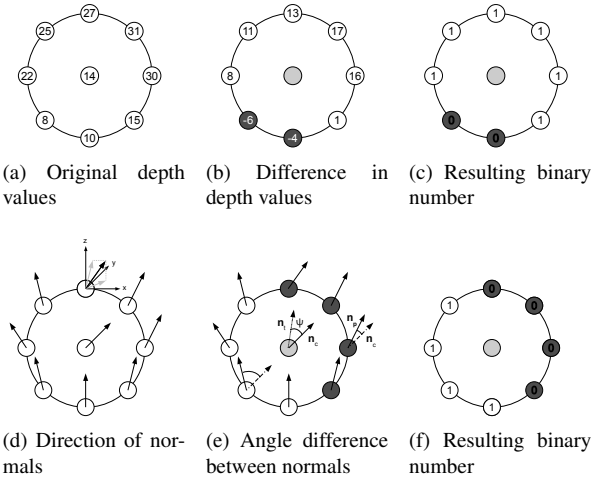
2. METHODOLOGY

In this section we describe the new LNBP feature descriptors, and the full method employed for facial action unit (AU) detection. In the preprocessing stage, the 3D facial geometries from the Bosphorus database [12] are first aligned via an affine transformation calculated from six of the provided facial feature points.

2.1. 3D Local Binary Patterns

3D Local Binary Patterns (3DLBPs) were proposed for use in facial expression recognition of 3D meshes in [10]. They exploit a 2D representation of the 3D information, the depth

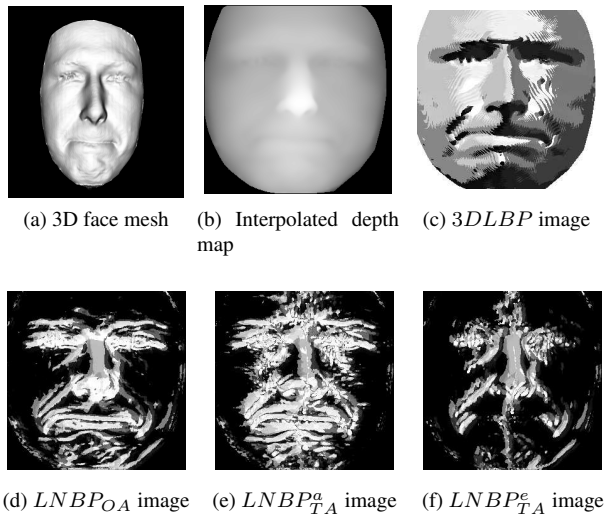
Work funded by European Research Council under the ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB). The work of G. Sandbach is funded by the Engineering and Physical Sciences Research Council through a DTA studentship. The work of S. Zafeiriou has been partially funded by Imperial College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



(a) Original depth values (b) Difference in depth values (c) Resulting binary number

(d) Direction of normals (e) Angle difference between normals (f) Resulting binary number

Figure 1: $3DLBP$ and $LNBP_{OA}$ operation using eight surrounding points. Figs. (a)-(c) show the $3DLBP$ operation, and Figs. (d)-(e) show the $LNBP_{OA}$ operation.



(a) 3D face mesh (b) Interpolated depth map (c) $3DLBP$ image

(d) $LNBP_{OA}$ image (e) $LNBP^{a}_{TA}$ image (f) $LNBP^{e}_{TA}$ image

Figure 2: Examples of images produced by the 3DLBP and LNBP feature descriptors for subject bs008 performing AU28.

map interpolated onto a regular grid, in order to encode the local shape around each point in the mesh. 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. This operation is shown in Figs. 1a-1c, and examples of a 3D facial mesh, depth map and resulting $3DLBP$ image for AU28 can be seen in Figs. 2a, 2b and 2c respectively.

2.2. Local Normal Binary Patterns

Local Normal Binary Patterns (LNBPs) employ the normals of the triangular polygons that form the 3D face mesh to encode the shape of the mesh at each point. This is equivalent to encoding the gradient of a 2D intensity image, and provides a richer source of information about the shape of the facial mesh than depth alone. Two novel feature descriptors are proposed in this paper, $LNBP_{OA}$ and $LNBP_{TA}$, both of which use the difference in the angles between normals to produce a binary number that encodes the shape of the neighbourhood around the central point, v_c.

As with LBPs, we define a circular neighbourhood around each point, specified by a radius r and P points regularly spaced around the circle. The unit normal, $\mathbf{n}_p$, at each point, v_p, in the neighbourhood is found, along with that at the central point, $\mathbf{n}_c$, through x-y interpolation of the given points in the mesh. The $LNBP_{OA}$ then calculates the scalar product of the two normals, and assigns a one or zero depending on whether this is higher or lower than the scalar product of the central normal with a vector that defined to be a given threshold angle, ψ, from the central normal, $\mathbf{n}_t$. This then produces a p-bit binary number for neighbourhood around each point in the grid. Alternatively, the $LNBP_{TA}$ calculates the difference in the two angles of the normals, the azimuth and the elevation, and assigns bits depending on how these compare to ψ_a and ψ_e respectively.

The $LNBP_{OA}$ is calculated as follows:

$$LNBP_{OA}(v_c) = \sum_{p=0}^{P-1} 2^p t(\langle \mathbf{n}_c, \mathbf{n}_p \rangle, \langle \mathbf{n}_c, \mathbf{n}_t \rangle) \qquad (1)$$

where $\langle \mathbf{n}_c, \mathbf{n}_t \rangle = cos(\psi)$ and

$$t(x, \tau) = \begin{cases} 1 & if \quad x < \tau \\ 0 & otherwise \end{cases} \qquad (2)$$

The two $LNBP_{TA}$ components are calculated as follows:

$$LNBP^{a}_{TA}(v_c) = \sum_{p=0}^{P-1} 2^p t(cos(|\theta_c - \theta_p|), cos(\psi_a)) \qquad (3)$$

$$LNBP^{e}_{TA}(v_c) = \sum_{p=0}^{P-1} 2^p t(cos(|\phi_c - \phi_p|), cos(\psi_e)) \qquad (4)$$

where $\theta = \arctan(\frac{y}{x})$ is the azimuth angle, and $\phi = \arctan(\frac{z}{x})$ is the elevation angle, of a normal $\mathbf{n} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.

The process of applying the $LNBP_{OA}$ feature descriptor is shown in Figs 1d-1f, and the $LNBP_{OA}$ image and the two $LNBP_{TA}$ images can be seen in Figs. 2d, 2e and 2f respectively. In these experiments 8 was used as the value of r and

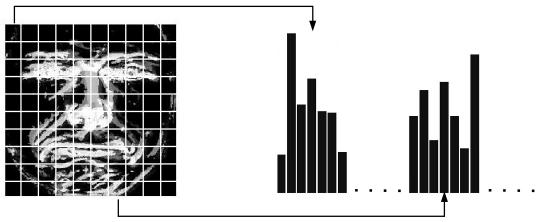


Figure 3: Construction of the feature vector by concatenation of histograms from each block in the image.

P, and $\frac{\pi}{12}$ was used for ψ, ψ_a and ψ_e. Different values of ψ were experimented with, and this value was found to generally give the best results.

2.3. Feature Vectors, Selection and Classification

Feature vectors are formed for each of the above descriptors through the use of histograms. First, the x-y plane of the mesh is divided into 10x10 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. This process can be seen in Fig. 3. In the case of the $LNBP_{TA}$ descriptor, a 2D histogram is formed from the two binary numbers produced, one from each angle. These histograms are patched together into a 2D image, and the result is then flattened into a 1D feature vector suitable for use with the SVMs. 60 bins were used to produce the 1D $3DLBP$ and $LNBP_{OA}$ feature descriptors, while 900 bins were employed to produce the 2D $LNBP_{TA}$ feature descriptor. In addition to exploiting the feature vector from a single descriptor, the benefits of the different descriptors were combined through feature fusion. This was achieved through concatenation of the two histogram feature vectors, thus allowing the classifier to choose features from either descriptor image.

Feature selection is performed in order to reduce the dimensionality of the feature vectors before classification. GentleBoost, a modified version of the AdaBoost algorithm, is used for this purpose. To avoid overfitting, our strategy is to run the selection algorithm repeatedly, removing previously chosen features at each stage, until the number of features selected exceeds the number of examples in the training set, or until fewer than 5 features are being chosen at each stage by the algorithm. This means that generally the number of features chosen will be about the same as the number of examples in the training set. SVM classifiers are trained for detection of each AU. These classifiers employed the histogram intersection kernel. Parameter optimisation was first performed using 5-fold cross-validation on a validation set which is taken as one third of the full training data, and then the SVMs are trained on the remaing two thirds.

3. EXPERIMENTAL RESULTS

We conducted experiments using the AU examples from the Bosphorus database [12], which consist of 3D facial data collected by asking 105 subjects to perform 24 of the facial AUs plus low intensity examples of AU12. In these experiments, all 24 AUs were employed for training and testing, with only the low intensity AU12 examples not included. As most subjects performed only a subset of the AUs, the subject order of the dataset was randomised before training and testing, to ensure that the AUs were more evenly distributed amongst the folds. 10-fold cross-validation was employed for training and testing the system. For training purposes, balanced sets were created for each AU by taking equal numbers of positive and negative examples, selecting the the negative examples randomly from the remaining AUs. Testing sets contained all of the positive and negative examples from the dataset. The ROC area under the curve (AuC) was calculated for each AU to assess the comparative performance of each feature type.

3.1. Single Descriptor Results

The result achieved when using each of the feature descriptors alone can be seen in the first three columns of Table 1 show the ROC AuC achieved for each of the AUs, with each of the 3DLBP and LNBP descriptors. These results demonstrate that on average all three descriptors achieve similar scores, with the LNBP descriptors achieving 95.17 ($LNBP_{OA}$) and 95.05 ($LNBP_{TA}$), just slightly lower than $3DLBP$ with 95.21. However, more interesting are the individual AU scores. Whilst for some AUs all three descriptors performed similarly, there are many for which the results varied significantly. The $3DLBP$ performs better than either LNBP descriptor for AUs 44, 9, 10, 16, 22, 23, 25, 27, and 34. However, there are a large number of AUs for which one or both of the LNBP descriptors outperforms the $3DLBP$: AUs 1, 4, 12, 12L, 12R, 14, 15, 17, and 20.

3.2. Feature Fusion Results

The ROC AuC achieved when the $3DLBP$ descriptor and each of the two LNBP descriptors were combined through feature fusion can be seen in the last two columns of Table 1. Due to the complementary nature of the single descriptor results, this was expected to result in a higher overall average performance, which is reflected by the improvement seen in the mean scores of 96.35 and 95.54 for $LBP{+}LNBP_{OA}$ and $LBP{+}LNBP_{TA}$ respectively, with a significant increase seen in the former case. However, looking at the individual AU scores, we can also see that for some AUs the result of combining two descriptors exceeds the maximum performance achieved with only one descriptor. For example, the results for AUs 2, 4, 15, 17, 24, 25 and 28 show that the AuC scores are higher when using $LBP{+}LNBP_{OA}$ than those achieved by either of the feature descriptors alone,

AU	$3DLBP$	BP1	BP2	BP3	BP4
1	92.60	96.37	92.80	**96.68**	91.05
2	97.93	97.94	97.43	**98.74**	98.43
4	95.81	96.05	96.58	**97.58**	96.86
43	99.35	99.24	99.65	99.34	**99.73**
44	**95.14**	88.79	90.38	94.83	94.12
9	**98.23**	96.80	97.69	97.21	97.92
10	**97.86**	96.87	95.74	97.85	97.70
12	95.58	96.37	96.32	**96.58**	95.96
12L	95.93	**98.86**	97.00	97.49	97.37
12R	97.72	**99.38**	97.21	97.69	97.40
14	91.38	92.88	92.13	**93.39**	91.39
15	84.95	90.15	87.00	**91.05**	83.61
16	**96.74**	93.46	95.14	96.36	96.68
17	93.39	95.53	95.03	**96.37**	94.68
18	96.83	96.19	96.82	**97.37**	97.25
20	90.22	92.94	92.71	**93.41**	92.51
22	**99.36**	97.95	98.30	99.31	99.32
23	94.60	90.82	93.08	94.20	**95.11**
24	88.90	89.06	88.15	**90.27**	89.63
25	93.37	92.52	92.42	**94.91**	94.82
26	93.00	93.39	**95.32**	94.37	95.08
27	**99.45**	97.90	99.07	99.38	99.38
28	97.59	97.70	97.08	**98.58**	97.95
34	99.01	97.01	98.02	**99.43**	99.11
μ	95.21	95.17	95.05	**96.35**	95.54
σ	0.54	0.82	0.45	**0.52**	0.46

Table 1: ROC AuC achieved with different features. BP1 - $LNBP_{OA}$, BP2 - $LNBP_{TA}$, BP3 - $3DLBP$+$LNBP_{OA}$, BP4 - $3DLBP$+$LNBP_{TA}$. μ - Mean AuC score across AUs, σ - Standard deviation of mean AuC over 10 folds.

which suggests that they are each allowing for different examples within these classes to be distinguished. For AUs 4 and 15 a significant improvement is seen over the 3DLBP result in this case. AUs 23 and 25 also demonstrate this extra improvement when the $3DLBP$ and $LNBP_{TA}$ descriptors are combined, with a significant improvement over 3DLBPs in the case of the latter AU. However, overall it would appear that the best results are achieved when the $3DLBP$ and $LNBP_{OA}$ feature descriptors are combined. Finally, we can compare the best result achieved with that of the work in [2], which uses curvature images with Gabor features to achieve a maximum ROC AuC score of 96.3 on AUs in this database, hence our fusion feature type compares well.

4. CONCLUSIONS

In this paper we proposed a new type of image feature descriptor, LNBPs, for use in encoding 3D facial geometry data for facial expression analysis. These descriptors were employed with GB feature selection and SVM classification in order to perform facial AU detection. The performance was compared to that achieved through use of the $3DLBP$ descriptor, and found to outperform this in the case of several AUs. The tests were then repeated with the two types of descriptors combined through feature fusion, and showed an overall improvement, particularly in the case of the $LNBP_{OA}$ when combined with the $3DLBP$. This suggests that the LNBP descriptors are able to encode shape information which is complementary to that encoded by 3DLBPs alone, and thus are beneficial for 3D facial AU detection.

5. REFERENCES

[1] Georgia Sandbach et al., "Recognition of 3D facial expression dynamics," *IVC*, 2012, in press.

[2] A. Savran et al., "Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units," *Pattern Recognition*, vol. 45, no. 2, pp. 767–782, 2012.

[3] C. Shan et al., "Facial expression recognition based on Local Binary Patterns: A comprehensive study," *IVC*, vol. 27, no. 6, pp. 803–816, 2009.

[4] S. Liao and A. Chung, "Face recognition with salient local gradient orientation binary patterns," in *ICIP*. IEEE, 2010, pp. 3317–3320.

[5] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," *Image and Signal Processing*, pp. 236–243, 2008.

[6] W. Zhang et al., "Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition," in *ICCV 2005*. IEEE, 2005, vol. 1, pp. 786–791.

[7] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE PAMI*, pp. 915–928, 2007.

[8] B. Jiang, M.F. Valstar, and M. Pantic, "Action unit detection using sparse appearance descriptors in space-time video volumes," in *FG 2011*, March 2011.

[9] S. Xie et al., "V-LGBP: Volume based local Gabor binary patterns for face representation and recognition," in *ICPR 2008*. IEEE, 2009, pp. 1–4.

[10] Y. Huang, Y. Wang, and T. Tan, "Combining statistics of geometrical and correlative features for 3d face recognition," in *BMVC*. Citeseer, 2006, pp. 879–888.

[11] D. Huang et al., "A novel geometric facial representation based on multi-scale extended local binary patterns," in *FG 2011*. IEEE, 2011, pp. 1–7.

[12] A. Savran et al., "Bosphorus database for 3D face analysis," *BIM*, pp. 47–56, 2008.