

CS-3DLBP and Geometry based Person Independent 3D Facial Action Unit Detection

Neslihan Bayramoğlu

nyalcinb@ee.oulu.fi

Guoying Zhao

gyzhao@ee.oulu.fi

Matti Pietikäinen

mkp@ee.oulu.fi

Center For Machine Vision Research, University of Oulu, Finland

Abstract

Face is the key component in understanding emotions which play significant roles in many areas from security and entertainment to psychology and education. In this paper, we propose a method to detect facial action units in 3D face data by combining novel geometric properties and a new descriptor based on the Local Binary Pattern (LBP) methodology. The proposed method enables person and gender independent facial action unit detection. The decision level fusion is used by employing the Random Forests classifiers to combine geometric and LBP based features. Unlike the previous methods which suffer from the diversity among different persons and normalize features utilizing neutral faces, our method extracts features on a single 3D face data. Besides, we show that orientation based 3D LBP descriptor can be implemented efficiently in terms of size and time without degrading the performance. We tested our method on the Bosphorus database and present comparative results with the existing methods. Our results outperform those of existing methods, achieving a mean receiver operating characteristic area under curve of 97.7%.

1. Introduction

Understanding emotions plays a significant role in social communications and many professions such as psychotherapy. Apart from these, there is a big demand in human-computer interaction research to analyse or transfer emotions to computer animated avatars and robots. Since emotions are primarily shown in face, it is the key component in understanding emotional expressions [6]. Therefore, facial expression analysis attracted many computer vision researchers into the field for years.

The majority of the previous works have focused primarily on the 2D data (images and videos) due to broad range of use, being readily available and because of the computational limitations. We refer reader to the comprehensive surveys in this area such as [16, 28] and [7].

This paper proposes a new set of geometric features and a modified Local Binary Patterns approach applied on the orientation of 3D face points for detecting 3D facial action units (AUs). We propose to combine these descriptors in the decision level by utilizing Random Forests classifiers [3].

1.1. Facial Action Coding Systems (FACS)

Most automatic expression analysis systems focus on recognizing a small set of emotional expressions, such as happiness, anger, surprise, disgust, sadness and fear [21, 27]. However, the frequency of occurring such prototypic expressions in everyday life conditions are infrequent relative to other facial expressions which are often composed of few facial feature movements [27]. Ekman and Friesen [5] proposed an anatomically based system for measuring facial movements. They claim to describe all visually distinguishable facial activity on the basis of 44 unique action units. Some of the AUs are shown in Figure 4. The system also allows coding the intensity of AUs on a five point scale. Facial expressions may contain single AU or combinations of several AUs. AUs can be considered as a basis spanning the facial expressions space by proper combinations. Therefore, automatic FACS AU recognition is better suited to analyse facial expressions rather than recognizing basic six emotions. However, due to subtle differences between certain AUs, automatic FACS AU detection problem is relatively difficult.

1.2. 3D Facial Action Unit Detection

Acquiring, processing and storing the 3D media increased rapidly during the last decade. 3D face recognition research also started to make use of these improvements; however, deeper understanding of 3D information for 3D facial expression recognition is still demanding. A recent survey [21] discusses available 3D facial expression recognition systems in detail. Existing approaches can be classified into two categories: feature-based and model-based. Majority of the methods for 3D facial expression recognition employs feature-based approaches [14, 19, 20, 22, 24,

26]. Distance based features are the most popular methods among feature-based methods [21]. Soyel and Demirel [24] employ six distances relying on the facial landmarks normalized with the width of the face. Similarly, Li et al. [14] use six distances normalized with the distance between eyes' inner points. Since faces of different people have different sizes and proportions, certain people may have similar head width and also similar eye separation but they may own different proportions in other facial parts (nose, mouth, etc.). Therefore, each facial feature should be normalized by its neutral state. This idea is employed in [26] by normalizing 96 distance based features with corresponding neutral states. Also, Srivasta and Roy [25] emphasize the drawbacks of employing absolute distances and instead utilize displacements between the positions of the facial points in neutral face and expression face. These distance based methods either suffer from the diversity among the faces of different people or require neutral faces which are not always available. In this study, instead of normalizing distances we propose using the ratios of distances and areas on a single 3D face data without needing neutral faces. We employ these features as geometrical features to recognize FACS AUs.

Most of the previous works employing static 3D facial expression data attempted to recognize six prototyping emotions instead of AUs [21]. This is mainly due to the structure of the available 3D facial expression datasets. There are few databases containing annotated FACS AUs such as Bosphorus database [22]. Savran et al. [22] utilized Bosphorus database to detect facial action units by mapping 3D facial surface geometry onto 2D. The 2D AU detection method with feature extraction and classification steps is then followed. Sandbach et al. [20] extended the LBP methodology to 3D face data by utilizing surface orientation information to detect AUs which is the closest work to our study. They later applied the traditional LBP operator to 2D representations of 3D face data [19]. In this study, we propose using Center Symmetric Local Binary Patterns (*CS-LBP*) [9] with the 3D surface orientation information. Compared to [20] our method produces shorter features leading to efficient computations and evaluations without degrading the performance.

The paper is organized as follows. Next section introduces the *CS-3DLBP* descriptor. Geometric facial features are then presented in the following section. Section 4 gives a brief summary of the Random Forests classifier. The comparative evaluation of our method with the state-of-the-art methods employing Bosphorus database is given in the results section.

2. CS-3DLBP

The well-known local binary descriptor ‘‘Local Binary Patterns’’ [15, 18] is shown to be quite successful in many

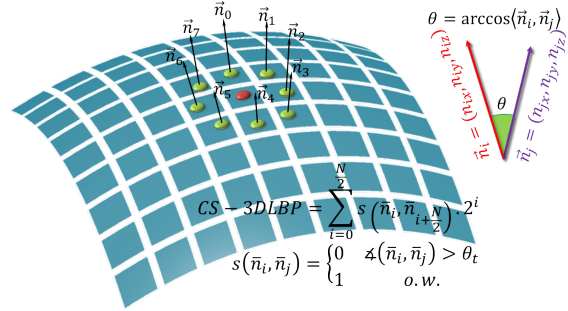


Figure 1. Construction of the orientation based *CS-3DLBP* operator on a surface defined on a regular grid.

computer vision problems particularly in face analysis research [1, 29]. Based on the LBP operator, Center Symmetric Local Binary Patterns (*CS-LBP*) [9] were developed for region description. *CS-LBP* produces smaller number of LBP labels thus results shorter histograms which are better suited for region description. Inspired from the *CS-LBP*, we propose a 3D local binary descriptor, namely Center Symmetric 3D Local Binary Descriptor (*CS-3DLBP*) which relies on the surface orientation information for facial action unit detection.

In literature, there are two main approaches for computing the surface normals. The classical method [11] estimates normal vector of a point by computing the tangent plane using k nearest neighbours. The least squares solution is utilized to obtain the plane parameters. In the second approach, polygon mesh structure is employed for estimating the orientation of a point (vertex). Orientations of the faces, which are incident to the vertex in question, are averaged to evaluate the normals [8]. A detailed survey on surface normal estimation methods can be found in [12]. In this study we utilize the triangular mesh structure of the face data and employ the latter method to obtain surface orientation.

In the original *CS-LBP* [9] method, which is an efficient region based image descriptor, intensity values of opposing pixels are compared and the differences are thresholded. In our case, where range image pixels correspond to 3D points, we compare the orientations, in terms of angle, between the opposing points to produce labels. The *CS-3DLBP* algorithm is illustrated in Figure 1. Initially, depth data are filtered for noise removal then normal vectors are estimated as explained previously. Angles between the normal vectors of opposing points are obtained by inner products (Equation 3). Similar to *CS-LBP*, our *CS-3DLBP* is defined as :

$$CS-3DLBP_{R,N,\theta_t}(p_c) = \sum_{i=0}^{N/2-1} s(\vec{n}_i, \vec{n}_{i+N/2}) 2^i \quad (1)$$

$$s(\vec{n}_i, \vec{n}_j) = \begin{cases} 1 & \angle(\vec{n}_i, \vec{n}_j) \geq \theta_t \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$\angle(\vec{n}_i, \vec{n}_j) = \arccos \frac{\langle \vec{n}_i, \vec{n}_j \rangle}{\|\vec{n}_i\| \cdot \|\vec{n}_j\|} \quad (3)$$

where $\vec{n}_i = (n_{ix}, n_{iy}, n_{iz})$ and $\vec{n}_{i+N/2}$ correspond to the surface normals at point pairs p_i and $p_{i+N/2}$ which are N equally spaced pixels on a circle of radius R centred at point $p_c = (p_{cx}, p_{cy}, p_{cz})$. The threshold function $s(\cdot)$ compares the angles between normal vectors of opposing points with a threshold θ_t . Since surface normals are unit vectors Equation 3 reduces to the numerator. Since the number of comparisons is halved compared to the LBP operator [20], CS-3DLBP produces $N/2$ -bit binary number resulting in $2^{N/2}$ distinct values for the binary pattern.

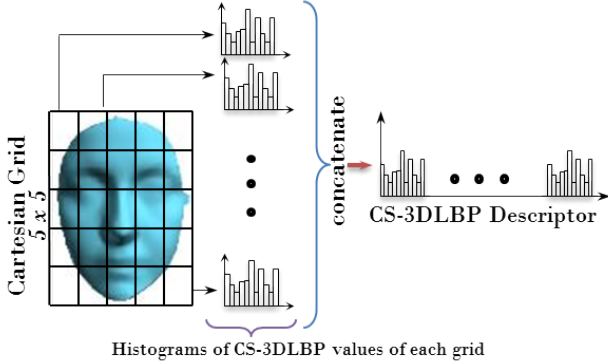


Figure 2. Construction of the *CS-3DLBP* descriptor.

To obtain the *CS-3DLBP* descriptor of a 3D face data, *CS-3DLBP* operator is applied to each surface point and the mapping utilizing these values is obtained. Figure 3d shows the *CS-3DLBP* mapped image of a 3D face data shown in Figure 3c. This mapping is then employed in constructing the *CS-3DLBP* descriptor employing a Cartesian grid. *CS-3DLBP* histograms are built on 5×5 grid (25 cells) and then concatenated into a single feature vector as shown in Figure 2.

3. Person Independent Geometric Features

Distance based features are the most popular methods for 3D facial expression recognition [21]. However these methods suffer from the diversity among different persons and among different ages (e.g. children vs. adult). To overcome this challenge, neutral faces are employed to specify the motion of feature points [14, 25] or to normalize [26] the features. However, neutral faces may not be available at all time. In this study, instead of using distances directly or normalizing these distances using neutral faces we propose using ratios of these distances and areas, and also angles defined on a single 3D face data. We believe that there is a harmony in every person’s face even if it has different size or shape than others. Therefore, in this study we tried to formulate the rules of this harmony for 3D expression face data.

A sample image from the Bosphorus database [22] and

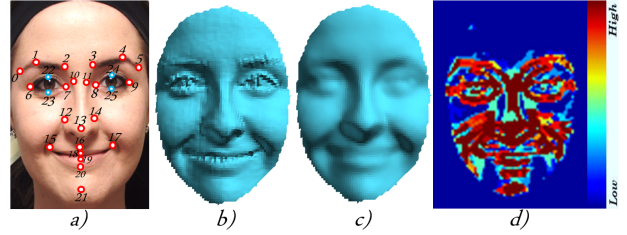


Figure 3. a) A sample 2D intensity data from Bosphorus database. Facial landmarks provided by the database (0-21) and points marked in this study (22-25) are shown with red and blue circles. b) Raw 3D data. c) Filtered 3D data and d) *CS-3DLBP* mapped image.

facial landmarks provided by the database (except eyelids) is shown in Figure 3a. Corresponding raw 3D model and its filtered output are shown in Figure 3b and 3c respectively. We propose using 24 features depending on the ratios of distances and areas, and the angles to detect facial action units. First seven of these features are presented in Figure 4. At the top of Figure 4, features are shown on a sample 3D face. The distance between two facial landmarks is denoted by d , areas of polygonal regions are shaded and denoted by A , and angles are denoted by α . Each feature is also associated with two samples from the database reflecting the significance of it. These samples and their AU names are shown just below the features. For example, first feature f_1 is the ratio of distance d_1 (left eye outer corner to mouth left corner) to the distance d_2 (right eye outer corner to mouth right corner). This ratio almost remains same for every person for the same expression. While it is a person independent feature, it is also a discriminative one especially for two specific lower face action units 12L and 12R. Similarly, other features that we employ in this study also exhibit discriminative person independent properties. Complete list of geometric features are listed in Table 1. In the table, features f'_i s are listed in the first column and associated formula is given in the second column. Facial landmarks, $p_i = (x_i, y_i, z_i) \in \mathbb{R}^3$, are denoted by letter p and landmark indexes i 's are consistent with the Figure 3. Distances d'_i s, areas A'_i s, and angles α'_i s are calculated in \mathbb{R}^3 . In the calculation of some features such as f_9 , mid points p'_i s of a line segment $\overline{p_i p_j}$ defined by points p_i and p_j are utilized. Triangular regions $\Delta(p_i p_j p_k)$ are formed by locating three landmark points p_i, p_j , and p_k as the corners of the triangles. Several regions and their corresponding areas are utilized as features. Angular features are calculated by employing vector inner products. Vectors $\overline{p_i p'_j}$ s are formed using two facial landmarks in which the vectors are pointing to p'_j s by convention. For obtaining the *ratio based geometric (RbG)* descriptor, all 24 features are concatenated into a single vector.

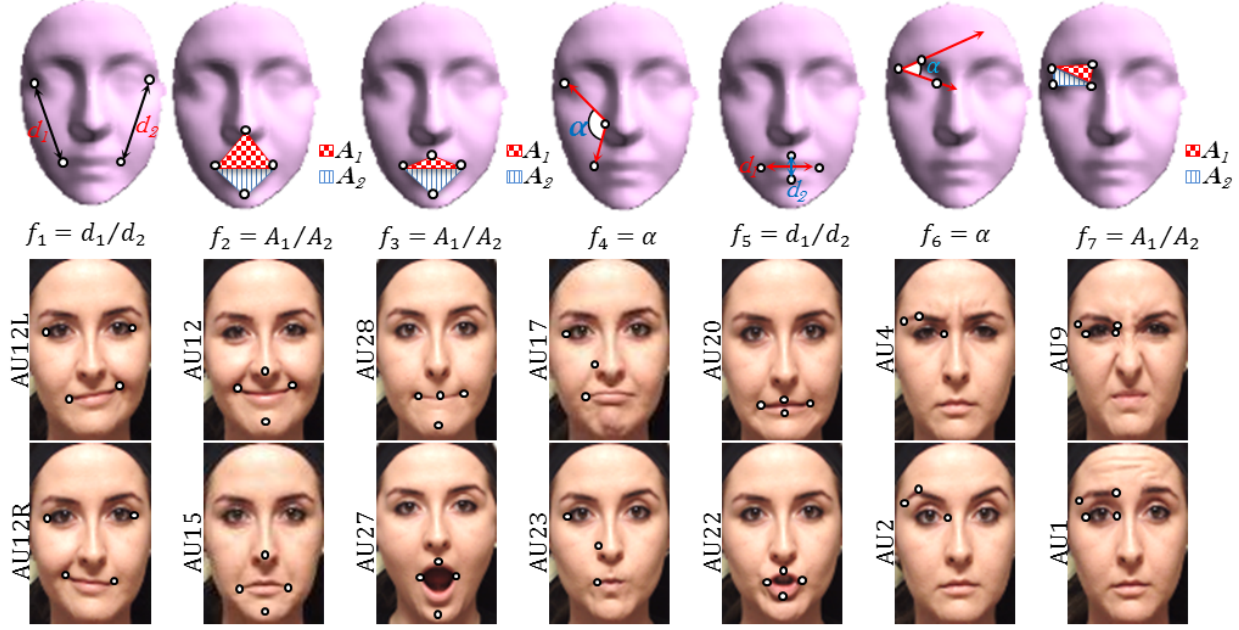


Figure 4. First seven person independent geometric features proposed in this study. First row demonstrates the construction of features. Associated equations are given underneath. Images shown in the last two rows are significant examples of the corresponding features with marked facial points used in computing features. 2D intensity images are shown for visibility reasons.

f_8	$d_1 = \ p_{10} - p_{11}\ ; d_2 = \ p_2 - p_3\ ; f = d_1/d_2$
f_9	$p_t = (p_{16} + p_{20}) \times 0.5; f = \angle(\overrightarrow{p_{15}p_t}, \overrightarrow{p_{17}p_t})$
f_{10}	$p_t = (p_{16} + p_{20}) \times 0.5; line_t = \overline{p_{15}p_{17}}$ $f = distance(p_t, line_t)$
f_{11}	$d_1 = \ p_2 - p_7\ ; d_2 = \ p_1 - p_7\ ; f = d_1/d_2$
f_{12}	$f = \angle(\overrightarrow{p_1p_2}, \overrightarrow{p_3p_4})$
f_{13}	$line_t = \overline{p_{15}p_{17}}; d_1 = distance(p_{16}, line_t)$ $d_2 = distance(p_{20}, line_t); f = d_1/d_2$
f_{14}	$d_1 = \ p_{13} - p_{20}\ ; d_2 = \ p_{13} - p_{21}\ ; f = d_1/d_2$
f_{15}	$f = \angle(\overrightarrow{p_{20}p_{15}}, \overrightarrow{p_{21}p_{15}})$
f_{16}	$line_t = \overline{p_{12}p_{14}}; line_k = \overline{p_{15}p_{17}}$ $d_1 = distance(p_{13}, line_t); d_2 = distance(p_{13}, line_k)$ $f = d_1/d_2$
f_{17}	$d_1 = \ p_{22} - p_{23}\ ; d_2 = \ p_6 - p_7\ ; f = d_1/d_2$
f_{18}	$d_1 = \ p_6 - p_7\ ; d_2 = \ p_{15} - p_{17}\ ; f = d_1/d_2$
f_{19}	$d_1 = \ p_{18} - p_{19}\ ; d_2 = \ p_{15} - p_{17}\ ; f = d_1/d_2$
f_{20}	$p_t = (p_{15} + p_{17}) \times 0.5; d_1 = \ p_{16} - p_t\ ;$ $d_2 = \ p_{20} - p_t\ ; f = d_1/d_2$
f_{21}	$d_1 = \ p_1 - p_{22}\ ; d_2 = \ p_1 - p_6\ ; f = d_1/d_2$
f_{22}	$f = \angle(\overrightarrow{p_{22}p_7}, \overrightarrow{p_{23}p_7})$
f_{23}	$d_1 = \ p_1 - p_{22}\ ; d_2 = \ p_1 - p_{23}\ ; f = d_1/d_2$
f_{24}	$A_1 = area\{\triangle(p_{22}p_6p_7)\}; A_2 = area\{\triangle(p_1p_6p_7)\}$ $f = A_1/A_2$

Table 1. Person Independent Geometric Features

4. Random Forests Classifier

Random Forests(RF) is an ensemble of tree predictors introduced by Ho [10] in 1995 and later studied in depth

by Breiman [3] in 2001. It is shown to be a fast and effective classification and regression method for many applications [13, 2, 23]. Each tree in the forest consists of split and leaf nodes. Features at each node are selected randomly and nodes are split into two (binary partitioning) by calculating the best split based on these randomly selected features. Best split is evaluated by maximizing the information gain of the split. Leaf nodes are created when the maximum tree depth is reached or the number of training samples at the node is less than the predefined threshold. During the training each leaf node may store the empirical class distributions associated to the subset of training data that has reached that leaf node [4]. During the test time, the query is sent down starting from the root node to the leaves through all trees in the forest. Finally, an ensemble class posterior is obtained by averaging all tree posteriors. Detailed information on decision forests can be found in [4]. The reason why we utilize RF classifiers instead of popular Support Vector Machines (SVM) classifiers is two fold. Firstly, we take the advantage of probabilistic outputs of RF models. Different classifiers having probabilistic outputs can easily be combined in the decision level to obtain a unified classifier. Secondly, the implicit feature selection in the RF model eliminates the need for another preprocessing step such as boosting which is employed in many previous studies [19, 20, 22]. We employ RF classifier as a detector (2 class problem). For each AU we evaluate *CS-3DLBP* descriptor and *RbG* descriptor separately and obtain two different class posterior probabilities which are used to decide

the presence or absence of AUs with a threshold probability of $p_{th} = 0.5$. Our combined descriptor (*CS-3DLBP+RbG*) is evaluated similarly by taking the mean value of the posteriors.

5. Experiments and Results

We tested our 3D AU detection methods on the Bosphorus database [22]. The database contains 105 subjects expressing up to 24 facial AUs resulting a total of 4666 face scans. The database also contains neutral faces, six prototypic emotions, occlusions, fixed rotations and corresponding intensity images. Some examples of facial action units are shown in Figure 4. Initially, 3D face data are normalized and depth maps are down-sampled to 75×100 (*width* \times *height*) images. Then a Gaussian filter is applied after a median filter to smooth the data (Figure 3c), because raw 3D data contains spikes and noise. Afterwards, mesh structure is obtained from the regularized point cloud representation in order to estimate the surface orientation as explained in Section 2.

The CS-3DLBP operator introduced in Section 2 has three parameters: radius R , number of neighbouring pixels N , and threshold on the angle difference of normals θ_t . The values 1 for R , 8 for N and 10 for θ_t are empirically found to yield good results. We used 16 bins to produce the CS-3DLBP feature descriptors for each equally sized grids obtained by dividing the image into 5×5 square regions. For computing some of our geometric features we manually marked eyelids p_{22} , p_{23} , p_{24} , and p_{25} on the 2D database images since these eyelid landmarks don't come along with the database. Although 24 landmarks are provided in the database we employ a small number of facial landmarks usually around 3 or 4 points. Some features can also be calculated by examining only one side of faces such as f_4 and f_6 . In this study we select the left side whenever it is available. For the Random Forests classifiers we limit tree depths to 5 and select the minimum number of samples at each node to be \sqrt{l} where l is the length of feature vectors. We used 2000 trees in RF model for training the detectors.

We compare our method with the two state-of-the-art methods [20] and [19]. Therefore, for evaluating our proposal we use the same experimental procedure in which cross-validation is used with 10 folds. In order to obtain person independent evaluation, subject based partitioning is employed. Subjects are randomly assigned to folds for obtaining balanced sets, since subjects in the database are performing different number of AUs. Detection performances are presented with the ROC area under curve (AuC) measure in Table 2 in a comparative manner. Detection results of our CS-3DLBP descriptor is given in the first column with an average AuC of 95.7%. Compared to [20] we can conclude that orientation based 3D LBP descrip-

		Method				
		CS-3DLBP	RbG	CS-3DLBP + RbG	Sandbach et al. [20]	Sandbach et al. [19]
Upper Face	Action Unit					
	AU1	93.4	93.6	96.5	96.4	95.7
	AU2	97.5	96.5	98.3	97.9	99.0
	AU4	95.2	90.0	94.9	96.1	97.9
	AU43	95.2	99.3	99.7	99.2	99.9
	AU44	91.1	90.3	93.6	88.8	95.7
	AU9	97.7	93.0	98.4	96.8	98.6
	AU10	98.5	91.8	98.1	96.9	97.6
	AU12	96.4	97.4	98.1	96.4	96.7
	AU12L	98.7	98.9	99.8	98.9	98.2
Lower Face	AU12R	97.3	99.0	99.4	99.4	98.3
	AU14	97.5	95.4	97.5	92.9	95.7
	AU15	92.8	94.7	96.6	90.2	92.7
	AU16	93.9	92.8	95.6	93.5	96.7
	AU17	94.7	98.8	98.6	95.5	96.9
	AU18	96.4	93.7	98.9	96.2	98.2
	AU20	92.9	98.1	95.8	92.9	95.1
	AU22	98.4	96.7	98.9	98.0	99.6
	AU23	95.5	92.2	98.3	90.8	96.4
	AU24	93.1	96.1	95.2	89.1	92.8
	AU25	91.2	96.1	96.8	92.5	95.4
	AU26	93.1	95.0	96.4	93.4	96.6
	AU27	98.9	97.9	99.7	97.9	99.7
	AU28	98.5	98.7	99.5	97.7	99.1
AU34	98.0	98.8	99.4	97.0	99.3	
Average		95.7	95.6	97.7	95.2	97.2

Table 2. ROC AuC (%) comparison.

tors can be implemented with shorter and faster way by utilizing CS-3DLBP descriptor without degrading the performance even with slight performance improvement. Similar to CS-3DLBP, our RbG descriptor has an average AuC of 95.6%. The reconstruction of RbG descriptor is simple, fast and doesn't rely too much on the performance of preprocessing steps. Besides the simplicity, its performance compares with the the state-of-the-art methods. For some action units such as AU20 and AU17, RbG descriptor comprises more discriminative features of the corresponding facial expressions. While dependency of our RbG descriptor on the facial landmarks seems to be a drawback of the method, there exists number of studies such as [17] for automatic landmark detection on the 3D face data. We combined our CS-3DLBP and RbG descriptors by averaging their posterior probabilities and achieved the highest detection performance. On the average a 97.7% AuC value is achieved which is slightly higher than the results of the best method reported in [19]. Particularly, our method is better in detecting lower facial action units and shows comparable performance for the upper facial action units.

6. Conclusions

In this study we showed that orientation based LBP descriptors can be computed efficiently by employing the *CS-LBP* strategy for 3D facial AU detection without degrading the performance. We also showed through experimentation that person independent geometric properties can be utilized for detecting 3D facial AUs without needing the neutral examples. We formulate the rules of geometric relations for AUs on a single expression face. This formulation achieves a considerably high AU detection rates for 3D face data and opens a promising direction for 2D facial action unit research. The quantitative results suggest that the combined method improves the overall detection performance, and achieves higher average ROC AuC for the Bosphorus database than the existing methods. However, we believe that there is a need for richer 3D face databases in terms of size and racial variability for evaluating and comparing the existing methods.

Acknowledgement: *The financial support from the Academy of Finland is gratefully acknowledged.*

References

- [1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *PAMI, IEEE*, 28(12):2037–2041, dec. 2006.
- [2] A. Bosch, A. Zisserman, and X. Muoz. Image classification using random forests and ferns. In *ICCV*, pages 1–8. IEEE, 2007.
- [3] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [4] S. J. Criminisi, A. and E. Konukoglu. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends® in Computer Graphics and Vision*, 7(2-3):81–227, 2011.
- [5] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement Consulting*. Consulting Psychologists Press, Palo Alto, 1978.
- [6] P. Ekman and W. Friesen. *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk, 2003.
- [7] B. Fasel and J. Luetttin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36(1):259–275, 2003.
- [8] H. Gouraud. Continuous shading of curved surfaces. *IEEE Transactions on Computers*, 100(6):623–629, 1971.
- [9] M. Heikkilä, M. Pietikäinen, and C. Schmid. Description of interest regions with local binary patterns. *Pattern Recognition*, 42(3):425–436, 2009.
- [10] T. Ho. Random decision forests. In *3rd Int. Conf. on Document Analysis and Recognition*, volume 1, pages 278–282. IEEE, 1995.
- [11] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle. *Surface reconstruction from unorganized points*, volume 26. ACM, 1992.
- [12] K. Klasing, D. Althoff, D. Wollherr, and M. Buss. Comparison of surface normal estimation methods for range sensing applications. In *ICRA*, pages 3206–3211. IEEE, 2009.
- [13] V. Lepetit, P. Laguerre, and P. Fua. Randomized trees for real-time keypoint recognition. In *CVPR*, pages 775–781, 2005.
- [14] X. Li, Q. Ruan, and Y. Ming. 3d facial expression recognition based on basic geometric features. In *10th Int. Conf. on Signal Pro. (ICSP)*, pages 1366–1369. IEEE, 2010.
- [15] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *PAMI, IEEE*, 24(7):971–987, jul 2002.
- [16] M. Pantic and L. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *PAMI, IEEE*, 22(12):1424–1445, 2000.
- [17] P. Perakis, G. Passalis, T. Theoharis, and I. Kakadiaris. 3d facial landmark detection under large yaw and expression variations. *PAMI, IEEE*, PP(99):1, 2012.
- [18] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen. *Computer Vision Using Local Binary Patterns*, volume 40. Springer, 2011.
- [19] G. Sandbach, S. Zafeiriou, and M. Pantic. Binary pattern analysis for 3d facial action unit detection. In *BMVC*, Guildford, UK, September 2012.
- [20] G. Sandbach, S. Zafeiriou, and M. Pantic. Local normal binary patterns for 3d facial action unit detection. In *IEEE, ICIP*, Orlando, FL, USA, October 2012.
- [21] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin. Static and dynamic 3d facial expression recognition: A comprehensive survey. *Image and Vision Computing*, 30(10):683–697, 2012.
- [22] 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.
- [23] J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, et al. Efficient human pose estimation from single depth images. *PAMI, IEEE*, 2012.
- [24] H. Soyel and H. Demirel. Facial expression recognition using 3d facial feature distances. *Image Analysis and Recognition*, pages 831–838, 2007.
- [25] R. Srivastava and S. Roy. 3d facial expression recognition using residues. In *IEEE Region 10 Conf., TENCON*, pages 1–5, 2009.
- [26] H. Tang and T. Huang. 3d facial expression recognition based on properties of line segments connecting facial feature points. In *8th Int. Conf. on Automatic Face & Gesture Recognition (FG)*, pages 1–6. IEEE, 2008.
- [27] Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. *PAMI, IEEE*, 23(2):97–115, 2001.
- [28] Z. Zeng, M. Pantic, G. Roisman, and T. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *PAMI, IEEE*, 31(1):39–58, 2009.
- [29] G. Zhao and M. Pietikäinen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *PAMI, IEEE*, 29(6):915–928, june 2007.