Eurographics Workshop on 3D Object Retrieval (2015)
I. Pratikakis, M. Spagnuolo, T. Theoharis, L. Van Gool, and R. Veltkamp (Editors)

A Spatio-Temporal Descriptor for Dynamic $3D$ Facial Expression Retrieval and Recognition

Antonios Danelakis †[1] and Theoharis Theoharis[1,2] and Ioannis Pratikakis[3]

[1]Department of Informatics & Telecommunications, University of Athens, Greece
[2]Department of Computer & Information Science, Norwegian University of Science and Technology, Norway
[3] Department of Electrical & Computer Engineering, Democritus University of Thrace, GR-67100, Xanthi, Greece

Abstract

The recent availability of dynamic 3D facial scans has spawned research activity in recognition based on such data. However, the problem of facial expression retrieval based on dynamic 3D facial data has hardly been addressed and is the subject of this paper. A novel descriptor is created, capturing the spatio-temporal deformation of the 3D facial mesh sequence. Experiments have been implemented using the standard $BU-4DFE$ *dataset. The obtained retrieval results exceed the state-of-the-art results and the new descriptor is much more frugal in terms of space requirements. Furthermore, a methodology which exploits the retrieval results, in order to achieve unsupervised dynamic 3D facial expression recognition is presented, in order to directly compare the proposed descriptor against the wealth of works in recognition. The aforementioned unsupervised methodology outperforms the supervised dynamic 3D facial expression recognition state-of-the-art techniques in terms of classification accuracy.*

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications—I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Curve, surface, solid, and object representations H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—Retrieval models

1. Introduction

Facial expressions are generated by facial muscle movements, resulting in temporary deformation of the face. In recent years, automatic analysis of facial expressions has emerged as an active research area due to its various applications such as human-computer interaction, human behavior understanding, biometrics, emotion recognition, computer graphics, driver fatigue detection, and psychology. Ekman [EF78] was the first to systematically study human facial expressions. His study categorizes the prototypical facial expressions, apart from neutral expression, into six classes representing anger, disgust, fear, happiness, sadness and surprise. This categorization is consistent across different ethnicities and cultures. Furthermore, each of the six aforementioned expressions is mapped to specific movements of facial muscles, called Action Units (*AU*s). This led to the Facial Action Coding System (*FACS*), where facial changes are described in terms of *AU*s.

The recent availability of $4D$ data† has increased research interest in the field. The first dataset that consists of $4D$ facial data was $BU-4DFE$, presented by Yin *et al.* [YCS*08]. $BU-4DFE$ was created at the University of New York at Binghamton and was made available in 2006. It involves 101 subjects (58 females and 43 males) of various ethnicities. For each subject the six basic expressions were recorded. The $Hi4D-ADSIP$ dataset was presented by Matuszewski *et al.* in [MQS*12]. The dataset was created at University of Central Lancashire and is not available yet. It contains 80 subjects (48 females and 32 males) of various age and

† This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES-3DOR (MIS 379516). Investing in knowledge Society through the European Social Fund.

† $4D$ will refer to $3D$ + time (dynamic $3D$); each element of such a sequence is a $3D$ frame.

DOI: 10.2312/3dor.20151056

ethnic origins. Each subject was recorded for seven basic expressions (anger, disgust, fear, happiness, sadness, surprise and pain). Finally, Yin *et al.* [ZYC*13] presented the $BP4D-Spontanous$ dataset in 2013 to the research community. This dataset contains high-resolution spontaneous $3D$ dynamic facial expressions. It involves 41 subjects (23 females and 18 males) of various ethnicities. Each of the aforementioned datasets are accompanied by a number of facial landmarks marked on each $3D$ frame. Table 1 illustrates the existing $4D$ facial expression datasets.

A lot of research has been dedicated to address the problem of facial expression recognition in dynamic sequences of $3D$ face scans. On the contrary, to the best of our knowledge, no much research on facial expression retrieval using dynamic $3D$ face scans appears in the literature. This paper illustrates results on the area of $4D$ facial expression retrieval. To this end, a novel descriptor is created, capturing the spatio-temporal deformation of the $3D$ facial mesh sequence. Preliminary experiments have been implemented using the standard dataset $BU-4DFE$. The obtained retrieval results are comparable to the state-of-the-art results but the new descriptor is much more flexible in terms of space complexity. Furthermore, a methodology which exploits the retrieval results, in order to achieve unsupervised dynamic $3D$ facial expression recognition, is presented. The aforementioned unsupervised methodology outperforms the supervised dynamic $3D$ facial expression recognition state-of-the-art techniques in terms of classification accuracy.

The remainder of the paper is organized as follows. In Section 2, previous works on the field of $4D$ facial expression retrieval are reviewed. In Section 3, the new spatio-temporal descriptor is explicitly described and the proposed retrieval methodology is illustrated. In Section 4, the experimental results of the proposed methodology are presented and discussed. Finally, conclusions and future challenges are drawn in Section 5.

2. Related Work

Due to the lack of previous work in $4D$ facial expression retrieval, the current section deals mainly with recognition; however, we concentrate on the descriptors and the $4D$ representation used, which are also related to the retrieval process. A detailed survey on $4D$ video facial expression recognition methodologies is presented in [DTP14b]. Methodologies are categorized based on the dynamic face analysis approach that they use. Dynamic face analysis enables robust detection of facial changes. Dynamic face analysis approaches can be divided into four categories: temporal tracking of facial landmarks, temporal tracking of facial critical points, mapping $3D$ facial scans onto a generic $3D$ face model and, finally, analyzing different facial surfaces in order to detect temporal facial changes.

2.1. Landmark Tracking-based Methods

Landmark tracking-based techniques aim to track areas around facial landmarks along $3D$ frames. Then, they detect temporal changes on geometry characteristics of the areas using appropriate features. The techniques presented in [CVTV05, RCY08, SCRY10, SRY08, SY08, TM09, TM10, CSZY12, DTP14a] belong to this category. In addition, the work presented in [DTP14a] is the only one dealing with $4D$ facial expression retrieval found in the literature. The proposed technique exploits eight facial landmarks in order to create the, so-called, $GeoTopo$ descriptor. $GeoTopo$ is a hybrid temporal descriptor which captures topological and geometric information of the $3D$ face scans along time.

2.2. Critical Point Tracking-based Methods

Critical points tracking-based techniques aim to track $3D$ model key points along $3D$ frames. Then, they detect temporal changes on spatial characteristics that are defined by these facial points and not by entire facial areas. The techniques presented in [BDBP12a, JLN*12] belong to this category.

2.3. 3D Facial Model-based Methods

Facial deformation-based techniques aim to generate descriptors based on the facial temporal deformations which occur due to facial expressions. To do so, they map each $3D$ facial scan onto a generic $3D$ face model and analyze the transformations taking place during the mapping. The techniques presented in [YWLB06, SZPR11, SZPR12, FZSK11, FZO*12, ZRY13] belong to this category.

2.4. Facial Surface-based Methods

Facial surface-based techniques extract facial surfaces on different face depth levels. The final descriptor is generated by estimating the intersection along time between the face and each surface. The techniques presented in [LTH11, DBAD*12] belong to this category.

3. Methodology

The large part of existing works on $4D$ facial expression analysis rely on facial landmarks/critical points, in order to build the corresponding descriptors. This happens because the $3D$ model-based dynamic face analysis approaches cannot operate reliably when pose variation is presented along the dynamic $3D$ sequence of the expression. In addition, facial expressions are closely linked to the positions of critical points of the face at given times. Furthermore, the development of the *FACS* [EF78], which describes the various facial movements in terms of AUs (see Figure 1), has not yet received the attention it deserves in the field of $4D$ facial expression analysis.

DATASET	YEAR	SIZE	CONTENT	LANDMARKS
$BU-4DFE$ [YCS*08]	2008	101 subjects	6 basic expressions	83 facial points
$Hi4D-ADSIP$ [MQS*12]	2012	80 subjects	7 basic expressions	84 facial points
$BP4D-Spontanous$ [ZYC*13]	2013	41 subjects	27 AUs	83 facial points

Table 1: *3D video facial expression datasets.*

The aforementioned reasoning led to the creation of the proposed descriptor for $4D$ facial expression retrieval. This new spatio-temporal descriptor captures the facial deformation along a dynamic $3D$ facial sequence. It is based on critical point-tracking face analysis. To this end, eight facial critical points are exploited for its creation (see Figure 2). The number of critical points used here is less than the number that is usually utilized by the state-of-the-art techniques and the algorithm for the detection of these eight points is founded on recent state-of-the-art work [PPTK13].

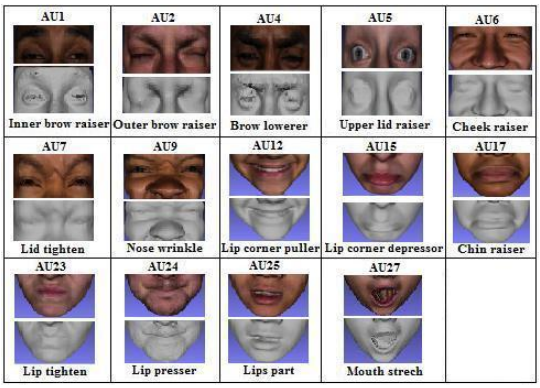

Figure 1: *The basic AUs as illustrated in Ekman's work.*

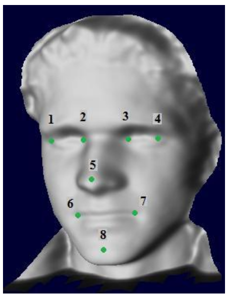

Figure 2: *The 8 facial critical points used for the creation of the proposed descriptor.*

3.1. The Proposed Descriptor

Each facial expression can be deconstructed into specific AUs, as illustrated in Table 2. There is a correspondence between each facial muscle and a number of AUs. The actual type of the AU is determined by the muscle temporal movement. For the creation of our descriptor we have chosen six features (i.e. two facial areas and four facial distances) and each one of them is directly related to one or more AUs of $FACS$, as illustrated in Table 3. $MEAN$ stands for the mean $3D$ point of two $3D$ points X, Y: $MEAN(X,Y) = \frac{X+Y}{2}$. The features have been selected in such a manner as to express the temporal motion of the AUs of the eyes, mouth and cheek. Moreover, according to the experimental results, these facial features are sufficient to distinguish the six expressions. The facial area formed by three $3D$ points is calculated using Heron's formula while the Euclidean formula is used for facial distances. Figures 3 and 4 illustrate the mapping of the selected six features onto a $3D$ face scan.

FACIAL EXPRESSION	ACTION UNITS
Anger	$\{AU4, AU7, AU23\}$
Disgust	$\{AU9, AU15\}$
Fear	$\{AU1, AU5, AU25\}$
Happiness	$\{AU6, AU12\}$
Sadness	$\{AU1, AU15, AU17, AU23\}$
Surprise	$\{AU1, AU5, AU26\}$

Table 2: *Facial expressions deconstruction into AUs.*

AU DESCRIPTION	FEATURE CODE	FEATURE TYPE	FEATURE VALUE
$AU6$: Cheek Raiser $AU17$: Chin Raiser	#1	Area	$\overbrace{CP1,CP5,CP6}^{AREA}$ or $\overbrace{CP4,CP5,CP7}^{AREA}$
$AU23$: Lip Tightener $AU25$: Lips Part	#2	Area	$\overbrace{CP6,CP7,CP8}^{AREA}$
$AU1$: Inner brow raiser $AU4$: Brow Lowerer $AU9$: Nose Wrinkle	#3	Distance	$\overline{MEAN(CP2,CP3),CP5}$
$AU12$: Lip Corner Puller $AU15$: Lip Corner Depressor	#4	Distance	$\overline{CP6,CP7}$
$AU5$: Lid Raiser $AU7$: Lid Tightener	#5	Distance	$\overline{MEAN(CP1,CP2),CP5}$ or $\overline{MEAN(CP3,CP4),CP5}$
$AU26$: Jaw Drop	#6	Distance	$\overline{CP1,CP8}$ or $\overline{CP4,CP8}$

Table 3: *Connecting AUs with mathematical features for the proposed descriptor.*

The proposed descriptor captures the facial deformation along the dynamic $3D$ facial sequence. To create the descriptor we use a $2D$ function (T), as illustrated in equation 1. Function T represents the value of the j-th feature, related to one or more AUs, in the i-th $3D$ frame. The calculations of the values of the aforementioned six features are performed using exclusively the $3D$ coordinates of the eight tracked critical points (CPs) on each $3D$ time frame. In other words, function T represents six different sequences of facial feature values for each dynamic $3D$ facial expression sequence.

$$T(i,j) = \left\{ \begin{array}{ll} Area_{i,j}(CPs) & : j \in \{1,2\} \\ Distance_{i,j}(CPs) & : j \in \{3,\ldots,6\} \end{array} \right\} \quad (1)$$

After the creation of T a subtraction scheme was implemented; the descriptor values are not used as absolute values corresponding to the current time frame, but as differences of the current from the initial time frame. To produce the final descriptor we apply the *Discrete Cosine Transformation* (DCT) on the subtracted spatial descriptor producing a transformed sequence for each feature. DCT represents a signal (or in our case a spatial sequence) as a sum of sinusoids of varying magnitudes and frequencies. It maps the features from the temporal to the frequency domain and thus the transformed features represent the spatio-temporal deformation of the initial features. Eight features of the transformed sequences are selected to construct the final descriptor. Equation 2 represents the final descriptor ST; the coefficients of DCT are real numbers and ST is an $8D$ vector irrespective of the number of frames of the corresponding facial expression $3D$ sequence.

$$ST = \left[\begin{array}{l} 2^{nd}\ DCT \text{ coefficient for area with feature code \#1,} \\ 3^{rd}\ DCT \text{ coefficient for area with feature code \#1,} \\ 3^{rd}\ DCT \text{ coefficient for area with feature code \#2,} \\ 2^{nd}\ DCT \text{ coefficient for distance with feature code \#3,} \\ 4^{th}\ DCT \text{ coefficient for distance with feature code \#3,} \\ \text{Mean of } DCT \text{ coefficients for distance with feature code \#4,} \\ 2^{nd}\ DCT \text{ coefficient for distance with feature code \#5,} \\ 2^{nd}\ DCT \text{ coefficient for distance with feature code \#6.} \end{array} \right] \quad (2)$$

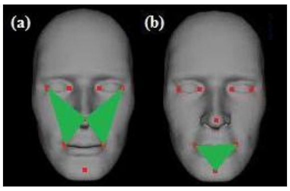

Figure 3: *Area features used for expressing (a) AU6 and AU17 (b) AU23 and AU25.*

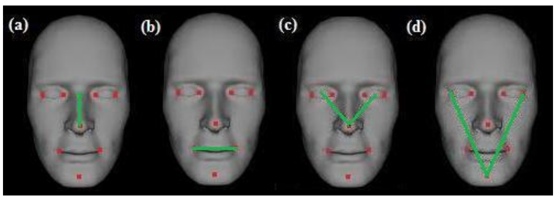

Figure 4: *Distance features used for expressing (a) AU1, AU4 and AU9 (b) AU12 and AU15 (c) AU5 and AU7 (d) AU26.*

Another transformation that could be used in order to map the features from the temporal to the frequency domain, and thus create spatio-temporal deformation of the initial features as well, is the *Fast Fourier Transformation* (FFT). FFT is similar to DCT, however, the experimental results proved that the implementation of DCT achieves much better results than FFT. This happens because DCT is much less complex than FFT, its coefficients are uncorrelated with each other and has better energy compaction [KTA*11]. This means that DCT has better ability than FFT to pack the information of the initial spatial sequence into as few frequency coefficients as possible.

For the comparison between two descriptors corresponding to different $4D$ data (query vs database descriptor), the *Kull-back Leibler Divergence* (KLD) [KL51] was implemented. The compared descriptors are of equal size, thus, KLD is extremely efficient. Given two descriptor vectors $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_N)$, where N is a positive integer, KLD yields optimal solution in $O(N)$ time. KLD is calculated using the formula $KLD = \sum_{i=1}^{i=N} \frac{x_i \cdot log(x_i)}{y_i}$, where *sum* represents the sum of the elements of the input vector. The closer to zero a returned KLD comparison value is, the more similar the two compared descriptors are, and thus, the more similar the two facial expressions.

4. Experimental Results

The dataset used to conduct our experiments is $BU-4DFE$. It was the first dataset consisting of faces recorded in $3D$ video, created by Yin *et al.* [YCS*08] at the University of New York at Binghamton. It was made available in 2006. It involves 101 subjects (58 females and 43 males) of various ethnicities. For each subject the six basic expressions were recorded. The faces were recorded gradually from neutral face, outset, apex, offset and back to neutral, using the dynamic facial acquisition system *Di*3*D* (www.di3d.com) and producing roughly 60,600 3D face models (frames), with corresponding texture images. The temporal resolution of the 3D videos is 25 fps and each 3D model consists of approximately 35,000 vertices. Finally, each frame is associated with 83 facial landmark points. In Figure 5, examples of $BU-4DFE$ dataset are illustrated.

The facial data constituting the dataset were preprocessed in order to be registered and of good quality. However, some inconsistencies are exhibited. Specifically, although in the database description [YCS*08], the authors state that each sequence contains an expression performed gradually from neutral appearance, low intensity, high intensity, and back to low intensity and neutral, this is not the case for some of the sequences (see Figure 6). Moreover, some videos contain corrupted meshes (see Figure 7) or they have obvious discontinuity. Finally, there are meshes that have spike shaped reconstruction artifacts around their borders. So, it is obvious that further improvement of the quality is a matter of significant importance. Berretti *et al.* [BDBP12b] presented a methodology in this direction, especially focusing on 3D static and dynamic facial data. It should be pointed out that, despite the aforementioned artifacts, no manual corrective removals took place.

Preliminary experiments have been conducted using the standard dataset $BU-4DFE$. Only the dynamic 3D sequences were used and not the corresponding textures. Six expressions for all 101 subjects of the dataset were used. Thus, over 60,600 3D frames were processed. In all tests, the *Leave-One-Out* approach was employed. In a pre-processing step, descriptor normalization took place, which sets the feature values of the descriptor in the interval $[0,1]$. Next, each feature of the proposed descriptor was weighed so that bigger weights correspond to features related to the facial areas around the mouth and eyes. The actual weights are given in Table 4. The weights were experimentally determined.

The experiments were divided in two groups. The first group involves experiments using three out of six expressions of the standard $BU-4DFE$ dataset, i.e. anger, happiness and surprise, similar to the approaches presented in [BDBP12a, SZPR11, DTP14a]. We did that in order for our method to be comparable with previous state-of-the-art approaches which have used only the aforementioned three expressions. The second group involves experiments using all six expressions provided by the standard $BU-4DFE$ dataset.

In Table 5 the retrieval evaluation metrics achieved by the new descriptor for three expressions are illustrated and compared to the only 4D facial expression retrieval technique found in the literature. Danelakis *et al.* [DTP14a] used three expressions of the $BU-4DFE$ dataset. We have used typical retrieval evaluation metrics such as Nearest Neighbor (NN), $1^{st}/2^{nd}$ tier and Discounted Cumulative Gain (DCG). In Figure 8 the corresponding precision-recall diagrams are presented. Our retrieval results for all six expressions of the standard $BU-4DFE$ dataset are illustrated in Table 6. These retrieval evaluation values are the first to be conducted on all six expressions of the $BU-4DFE$ dataset and are very promising.

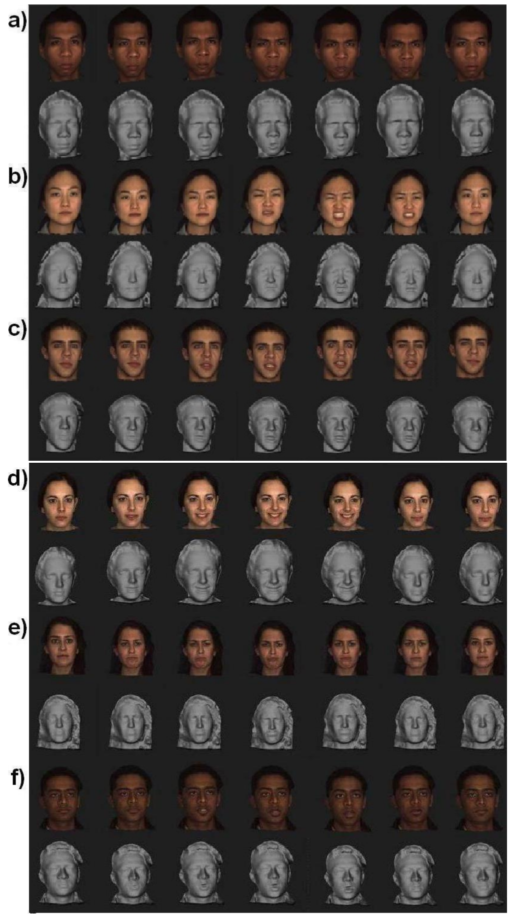

Figure 5: *Example of $BU-4DFE$ dataset including texture images and 3D models: (a) anger, (b) disgust, (c) fear, (d) happiness, (e) sadness and (f) surprise.*

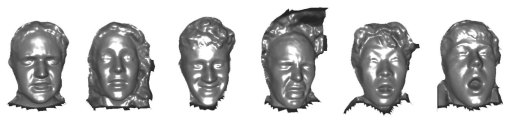

Figure 6: *Initial frames from $BU-4DFE$ dataset sequences in which the subjects do not start with a neutral expression.*

The proposed spatio-temporal descriptor can be used to implement 4D facial expression recognition as well. This

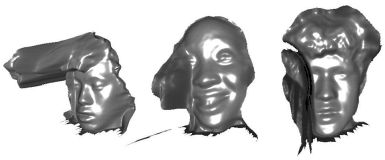

Figure 7: *Illustration of corrupted frames in the $BU-4DFE$ dataset.*

FEATURE	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}
WEIGHTS	0.10	0.20	0.10	0.10	0.20	0.10	0.10	0.10

Table 4: *Feature weights in the proposed descriptor.*

allows our method to be compared against state-of-the-art methods whose performance is evaluated in terms of classification accuracy. Compared to the existing approaches, the process illustrated here is completely unsupervised but is better in terms of classification accuracy. To achieve 4*D* facial expression recognition, by exploiting the 4*D* facial retrieval results of the proposed descriptor, a majority voting is implemented among the *k*-top retrieval results. The query 4*D* facial expression is classiffied as belonging to the outvoting class within the *k*-top retrieved results. In Table 7 the classification accuracies achieved by our descriptor, with respect to the variable *k*, are outlined.

Table 8 summarizes the performance of state-of-the-art methods on 4*D* facial expression recognition for the expressions from the $BU-4DFE$ dataset. It should be pointed out that landmark-based techniques of Table 8 use their own automatic procedure to detect facial 3*D* landmarks. Furthermore, Danelakis *et al.* use the landmarks provided by $BU-4DFE$ dataset. In addition, Le *et al.*'s [LTH11] method uses the sad instead of angry expression. Danelakis *et al.* [DTP14a] and the proposed work achieve completely unsupervised recognition. On the other hand, the rest of the methods presented in Table 8 use subsets of $BU-4DFE$, as training sets, in order to train their implemented classifiers.

5. Conclusions and Future Work

Dynamic 3*D* facial expression analysis constitutes a crucial open research field due to its applications in human-computer interaction, psychology, biometrics etc. In this paper, a new approach for dynamic 3*D* facial expression retrieval is presented and a novel spatio-temporal descriptor is proposed. This descriptor captures the facial expression deformation of 3*D* face scans along time. Preliminary experiments have been conducted on the standard dataset $BU-4DFE$. The obtained results are very promising and can be provided as ground truth for future retrieval techniques. Furthermore, a methodology which exploits the retrieval results, in order to achieve unsupervised dynamic 3*D* facial expression recognition, is presented. This methodology achieves better classification accuracy than the super-

METHOD	*NN*	1^{st} *TIER*	2^{nd} *TIER*	*DCG*
Danelakis *et al.* [DTP14a]	0.88	0.74	0.90	0.89
Proposed Method	**0.88**	**0.76**	**0.94**	**0.94**

Table 5: *Retrieval evaluation for the proposed descriptor on $BU-4DFE$ (3 expressions).*

METHOD	*NN*	1^{st} *TIER*	2^{nd} *TIER*	*DCG*
Proposed Method	**0.75**	**0.61**	**0.66**	**0.86**

Table 6: *Retrieval evaluation for the proposed descriptor on $BU-4DFE$ (6 expressions).*

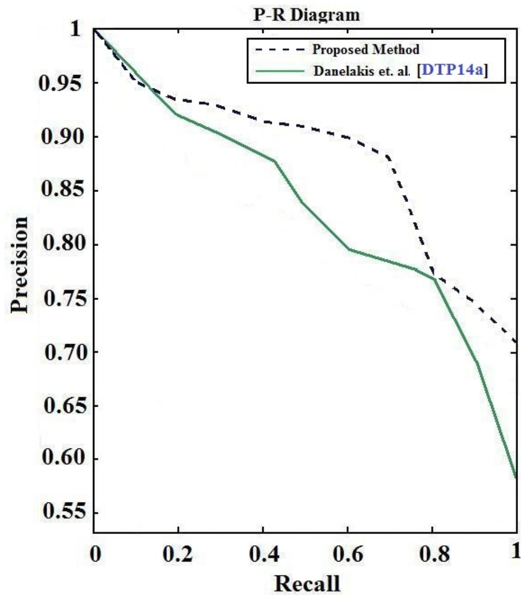

Figure 8: *Precision-Recall diagram for the proposed descriptor on $BU-4DFE$ (3 expressions).*

k	CLASSIFICATION ACCURACY (%)
3	90.83
5	85.53
10	78.20
15	73.53
20	73.53
50	73.53
100	73.72

Table 7: *Classification accuracies of the proposed descriptor on $BU-4DFE$ (6 expressions).*

METHOD	NUMBER OF EXPRESSIONS	CLASSIFIER TRAINING	CLASSIFICATION ACCURACY
Sun *et al.* [SCRY10]	6	YES	94.37%
Drira *et al.* [DBAD*12]	6	YES	93.21%
Fang *et al.* [FZO*12]	6	YES	91.00%
Proposed Method	**6**	**NO**	**90.83%**
Canavan *et al.* [CSZY12]	6	YES	84.80%
Berretti *et al.* [BDBP13]	6	YES	79.40%
Jeni *et al.* [JLN*12]	6	YES	78.18%
Zhang *et al.* [ZRY13]	6	YES	76.12%
Fang *et al.* [FZSK11]	6	YES	75.82%
Sandbach *et al.* [SZPR12]	6	YES	64.60%
Proposed Method	**3**	**NO**	**99.67%**
Danelakis *et al.* [DTP14a]	3	NO	96.67%
Le et al. [LTH11]	*3*	YES	*92.22%*
Sandbach *et al.* [SZPR11]	3	YES	81.93%
Berretti *et al.* [BDBP12a]	3	YES	76.30%

Table 8: *Evaluation of the proposed descriptor against the state-of-the-art on dynamic 3D facial expression recognition using the $BU-4DFE$ dataset.*

vised dynamic $3D$ facial expression recognition state-of-the-art techniques.

The further improvement of the $3D$ landmark detection algorithm [PPTK13] is an issue that will be addressed in the future. At present, the detection algorithm is executed separately for each $3D$ frame of the time sequence. The aim is to exploit the $3D$ positions of the critical points of the previous frame in order to find the corresponding points in the current time frame. This performance improvement can lead to real-time implementation. In addition, the proposed methodology will be extended to handle all the remaining expressions of $BU-4DFE$ dataset. Arbitrary expressions will also be taken into account.

References

[BDBP12a] BERRETTI S., DEL BIMBO A., PALA P.: Real-time expression recognition from dynamic sequences of 3D facial scans. In *EG Workshop on 3D Object Retrieval* (2012), pp. 85–92. 2, 5, 7

[BDBP12b] BERRETTI S., DEL BIMBO A., PALA P.: Super-faces: A super-resolution model for 3D faces. In *Computer Vision – ECCV 2012. Workshops and Demonstrations*, vol. 7583. Springer Berlin Heidelberg, 2012, pp. 73–82. 5

[BDBP13] BERRETTI S., DEL BIMBO A., PALA P.: Automatic facial expression recognition in real-time from dynamic sequences of 3D face scans. *Vis. Comput. 29*, 12 (2013), 1333–1350. 7

[CSZY12] CANAVAN S. J., SUN Y., ZHANG X., YIN L.: A dynamic curvature based approach for facial activity analysis in 3D space. In *CVPR Workshops* (2012), pp. 14–19. 2, 7

[CVTV05] CHANG Y., VIEIRA M. B., TURK M., VELHO L.: Automatic 3D facial expression analysis in videos. In *IEEE Workshop AMFG '05* (2005), pp. 293–307. 2

[DBAD*12] DRIRA H., BEN AMOR B., DAOUDI M., SRIVASTAVA A., BERRETTI S.: 3D dynamic expression recognition based on a novel deformation vector field and random forest. In *ICPR '12* (2012), pp. 1104–1107. 2, 7

[DTP14a] DANELAKIS A., THEOHARIS T., PRATIKAKIS I.: Geotopo: Dynamic 3D facial expression retrieval using topological and geometric information. In *Proc. 3D Object Retrieval 2014* (2014), pp. 1–8. 2, 5, 6, 7

[DTP14b] DANELAKIS A., THEOHARIS T., PRATIKAKIS I.: A survey on facial expression recognition in 3D video sequences. *Multimedia Tools and Applications* (2014), 1–39. 2

[EF78] EKMAN P., FRIESEN W.: *Facial action coding system: A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto, 1978. 1, 2

[FZO*12] FANG T., ZHAO X., OCEGUEDA O., SHAH S. K., KAKADIARIS I. A.: 3D/4D facial expression analysis: An advanced annotated face model approach. *Image and Vision Computing 30*, 10 (2012), 738–749. 2, 7

[FZSK11] FANG T., ZHAO X., SHAH S. K., KAKADIARIS I. A.: 4D facial expression recognition. In *ICCV '11* (2011), pp. 1594–1601. 2, 7

[JLN*12] JENI L. A., LÓRINCZ A., NAGY T., PALOTAI Z., SEBÓK J., SZABÓ Z., TAKÁCS D.: 3D shape estimation in video sequences provides high precision evaluation of facial expressions. *Image and Vision Computing 30*, 10 (2012), 785 – 795. 2, 7

[KL51] KULLBACK S., LEIBLER R. A.: On information and sufficiency. *Annals of Mathematical Statistics 22* (1951), 49–86. 4

[KTA*11] KEKRE H., THEPADE S., ATHAWALE A., SHAH A., VERLEKAR P., SHIRKE S.: Performance evaluation of image retrieval using energy compaction and imagetiling over DCT row mean and DCT column mean. In *Thinkquest 2010*, Pise S., (Ed.). Springer India, 2011, pp. 158–167. 4

[LTH11] LE V., TANG H., HUANG T. S.: Expression recognition from 3D dynamic faces using robust spatio-temporal shape features. In *IEEE FG '11* (2011), pp. 414–421. 2, 6, 7

[MQS*12] MATUSZEWSKI B., QUAN W., SHARK L., MCLOUGHLIN A., LIGHTBODY C., EMSLEY H., WATKINS C.: Hi4D-ADSIP 3D dynamic facial articulation database. *Elsevier Image and Vision Computing 30*, 10 (2012), 713–727. 1, 3

[PPTK13] PERAKIS P., PASSALIS G., THEOHARIS T., KAKADIARIS I. A.: 3D facial landmark detection under large yaw and expression variations. *IEEE Transactions on Pattern Analysis and Machine Intelligence 35*, 7 (2013), 1552–1564. 3, 7

[RCY08] ROSATO M., CHEN X., YIN L.: Automatic registration of vertex correspondences for 3D facial expression analysis. In *IEEE International Conference on Biometrics: Theory, Applications and Systems* (2008), pp. 1–7. 2

[SCRY10] SUN Y., CHEN X., ROSATO M. J., YIN L.: Tracking vertex flow and model adaptation for three-dimensional spatiotemporal face analysis. *IEEE Transactions on Systems, Man, and Cybernetics, Part A 40*, 3 (2010), 461–474. 2, 7

[SRY08] SUN Y., REALE M., YIN L.: Recognizing partial facial action units based on 3D dynamic range data for facial expression recognition. In *FG '08* (2008), pp. 1–8. 2

[SY08] SUN Y., YIN L.: Facial expression recognition based on 3D dynamic range model sequences. In *Springer Proc. ECCV '08: Part II* (2008), pp. 58–71. 2

[SZPR11] SANDBACH G., ZAFEIRIOU S., PANTIC M., RUECKERT D.: A dynamic approach to the recognition of 3D facial expressions and their temporal models. In *IEEE FG '11* (2011), pp. 406–413. 2, 5, 7

[SZPR12] SANDBACH G., ZAFEIRIOU S., PANTIC M., RUECKERT D.: Recognition of 3D facial expression dynamics. *Elsevier Image and Vision Computing 30*, 10 (2012), 762–773. 2, 7

[TM09] TSALAKANIDOU F., MALASSIOTIS S.: Robust facial action recognition from real-time 3D streams. In *CVPR '09* (2009), pp. 4–11. 2

[TM10] TSALAKANIDOU F., MALASSIOTIS S.: Real-time 2D+3D facial action and expression recognition. *Elsevier Pattern Recognition 43*, 5 (2010), 1763–1775. 2

[YCS*08] YIN L., CHEN X., SUN Y., WORM T., REALE M.: A high-resolution 3D dynamic facial expression database. In *IEEE Proc. FG '08* (2008), pp. 1–6. 1, 3, 4, 5

[YWLB06] YIN L., WEI X., LONGO P., BHUVANESH A.: Analyzing facial expressions using intensity-variant 3D data for human computer interaction. In *Proc. ICPR '06* (2006), pp. 1248–1251. 2

[ZRY13] ZHANG X., REALE M., YIN L.: Nebula feature: A space-time feature for posed and spontaneous 4D facial behavior analysis. In *IEEE FG '13* (2013). 2, 7

[ZYC*13] ZHANG X., YIN L., COHN J. F., CANAVAN S., REALE M., HOROWITZ A., LIU P.: A high-resolution spontaneous 3D dynamic facial expression database. In *IEEE FG '13* (2013). 2, 3