Facial Action Unit Detection: 3D versus 2D Modality

Arman Savran
Electrical and Electronics Engineering
Boğaziçi University, Istanbul, Turkey
arman.savran@boun.edu.tr

Bülent Sankur
Electrical and Electronics Engineering
Boğaziçi University, Istanbul, Turkey
bulent.sankur@boun.edu.tr

M. Taha Bilge
Department of Psychology
Boğaziçi University, Istanbul, Turkey
taha.bilge@boun.edu.tr

Abstract

In human facial behavioral analysis, Action Unit (AU) coding is a powerful instrument to cope with the diversity of facial expressions. Almost all of the work in the literature for facial action recognition is based on 2D camera images. Given the performance limitations in AU detection with 2D data, 3D facial surface information appears as a viable alternative. 3D systems capture true facial surface data and are less disturbed by illumination and head pose. In this paper we extensively compare the use of 3D modality vis-à-vis 2D imaging modality for AU recognition. Surface data is converted into curvature data and mapped into 2D so that both modalities can be compared on a fair ground. Since the approach is totally data-driven, possible bias due to the design is avoided. Our experiments cover 25 AUs and is based on the comparison of Receiver Operating Characteristic (ROC) curves. We demonstrate that in general 3D data performs better, especially for lower face AUs. Furthermore it is more robust in detecting low intensity AUs. Also, we show that generative and discriminative classifiers perform on a par with 3D data. Finally, we evaluate fusion of the two modalities. The highest detection rate was achieved by fusion, which is 97.1 area under the ROC curve. This score was 95.4 for 3D and 93.5 for 2D modality.

1. Introduction

Mental processes like emotions, social interactions as in winking and physiological effects like pain or fatigue do all generate expressions on human faces, consciously or unconsciously. From behavioral scientists to human-computer interaction analysts there is a need for objective facial measurements. Facial Action Coding System (FACS) seems to provide the right medium [2, 11]. The downside of FACS is that due to the countless rules defined in FACS and subtleties of AUs, encoding is a difficult and time consuming process, and requires certification. Also, due to the human factor in coding, analysis can be subjective by nature. These are the reasons that started vision based automated AU coding research. The potential of an automatic AU encoder is obvious in the design of human-computer interfaces or in psychological studies, to infer social communication facial signals, mental and physiological activities.

An approach for AU detection is to develop specific models and to identify relevant image features. For instance, Tian *et al*. [16] fits geometric models of facial features in the form of ellipses and curves, and applies wrinkle detectors. Pantic and Rothcrantz [8] extract geometric features from fiducial points, and they use profile as well as frontal views to benefit from depth information. Statistical face models, like active shape models (ASM) and active appearance models (AAM) are proven techniques for model based facial analysis. Lucey *et al*. [5] apply both 2D and 3D AAM tracking to classify four upper face AUs.

The performance of model-based techniques can critically depend upon the success of intermediate steps like landmark detection and localization. In contrast, data-driven techniques, i.e., direct image-based analysis has the double advantage of avoiding the weakest-link-in-the-chain issue of the intermediate steps, and also their training phase is relatively simpler. Bartlett *et al*. [1] proposed a successful image-based technique which works in a totally data-driven fashion. Their method performs local analyses by Gabor wavelets whose coefficients are chosen automatically with AdaBoost algorithm.

Despite considerable progress on the automatic detection of AUs using 2D image data, several challenges remain, for example, detection performance with subtle AUs or when several AUs co-occur in combinations. It is conjectured that 3D face data can alleviate difficulties inherent in 2D modality since 3D enables true facial surface measurements. The advantages of 3D measurements have already been demonstrated in the context of face recognition, as 3D is immune to illumination and to some extent to pose variations. The potential of AU detection with 3D observations has not been sufficiently explored, though there has been some work done on 3D emotions (the six universal ones [3], happiness, surprise, fear, anger, sadness, disgust). For instance, using 64 manually marked points, Wang *et al.* [18] divide 3D faces into regions and extract regional histograms of surface curvatures. Other 3D methods proposed for emotion identification have the handicap of depending on an excessive number of feature points [12, 6, 7, 13, 14].

Use of simultaneous 2D and 3D video acquired by structured light based system has been proposed for AU recognition by Tsalakanidou and Malassiotis [17]. They employ ASM based tracker, and also utilize detectors of eyebrows and mouth for a more robust ASM fitting; and then apply rule-based classification on features related to facial point distances, gray-level appearance and surface curvatures. However, their experimentation is limited to 11 singly occurring posed AUs and the performance results seem lower than current 2D AU detection. In our previous study we [10] we used solely 3D data and introduced a non-rigid registration-based approach for surface curvature images. The resulting ROC analysis revealed 96.2% Area under the Curve (AuC) score over 22 singly occurring AUs.

A relevant problem to be explored is whether 2D imaging modality or 3D modality capturing surface data is better and whether they can be complementary. With this goal in mind we extensively compare 3D and 2D modalities for 25 AUs over very comprehensive 3D FACS dataset [9]. The 2D camera images were analyzed with the state-of-the-art method of Bartlett *et al.* [1]. It is also important to remark that for the sake of a fair comparison, the 3D face data was converted into surface curvature data, and then mapped into 2D. This is to minimize any confounding factors so that identical experimental conditions are created for the 2D to 3D comparison. Therefore in the experiments only the data modalities differ, but not the methods.

The paper is organized as follows. Detection from 2D images, involving generative in addition to discriminative classification, is described in Section 2. Section 3 explains use of 3D data. In section 4 we give our 3D-2D modality comparison results by discussing their differences on per AU basis, examine the case of low intensity AUs and evaluate fusion of the two modalities. Conclusion are drawn and hints for future work are given in Section 5.

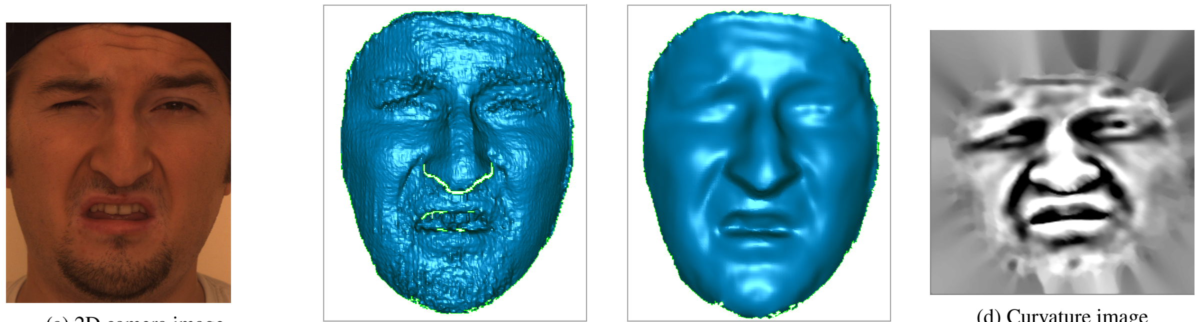
2. Action Unit Detection from 2D Images

In order to detect AUs from 2D images the method proposed by Bartlett *et al.* [1] is employed. This is the state-of-art-method for automatic AU recognition in 2D. Since the method is data-driven, hence in a sense free from biases of model-driven methods, it is a natural platform to compare the performances of the 2D and the 3D modalities. More explicitly, the outcome of the comparison does not depend on the performance of intermediate steps like model fitting or facial landmark detection. In this method, facial features are analyzed by Gabor wavelets. Since Gabor basis effectively support only limited part of the input images, especially more compact regions for higher frequencies, the analysis is local in contrast to holistic methods like PCA where basis cover all the image pixels. In the training phase, AdaBoost feature selection is applied on the Gabor magnitude responses for each AU separately. Although the resulting AdaBoost classifiers could have been used as detectors, it is shown that linear Support Vector Machine (SVM) classifiers on the selected features improve the performance.

Several other works in the literature consider discriminative classifiers like AdaBoost, SVMs or neural networks. Despite certain success achieved by discriminative classifiers they have some weaknesses. Their generalization capability may be poor when training samples do not adequately represent all possible variations. Also, inaccuracies of the ground-truth labeling may have more severe impact on the discriminative models. Both issues are of real concern in AU detection since not all AUs are richly represented in the present databases, and more importantly, FACS coding is sometimes subjective, with acceptance of approximately 75% reported agreement between annotators [2, 11]. Therefore, we opt for generative classifiers. Actually, since classification is carried out on the most discriminative features selected by AdaBoost, both discriminative and generative characteristics exist. Four types of Bayes classifiers are tested for this purpose, where features are assumed to be Gaussian:

- **G-1:** Quadratic Normal classifier with diagonal covariance, where the covariance matrices are estimated for positive and negative samples separately;
- **G-2:** Simplest Quadratic classifier, where single global variance is estimated for each class.
- **G-3:** Linear Normal classifier with diagonal covariance (Naïf Bayes).
- **G-4:** Simplest Linear Classifier (Nearest Mean class.).

In order to eliminate any confounding effect due to automatic normalization, images have been normalized using manually determined eye centers. Normalization involves



(a) 2D camera image (b) Raw 3D face (c) Filtered 3D face (d) Curvature image

Figure 1: Illustration of pre-processing steps of 3D surface data to produce surface curvature images. Input data (b) is filtered (c) and its surface curvature is computed. The curvature information is projected onto 2D domain and extrapolation is performed (d). The FACS code for this particular expression is R1C+L2D+4B+R6B+G7E+10E+19B+25C+26C+43B, which includes many lower and upper facial AUs, and also some asymmetry is involved.

translation, rotation and scaling to align eye centers and resampling of images at 96×96 pixels resolution. For Gabor transform analysis of faces, eight directions are used and the Gabor wavelengths are chosen in the range of 2-32 pixels, which vary with half octave intervals, and result in nine scale levels. The feature vector has $9 \times 8 \times 96 \times 96 = 663552$ components; however, not all of them are informative and in fact for any one AU, 200 Gabor features selected by means of AdaBoost, seem to suffice.

3. From 3D Surface Measurements to 2D Curvature Field

Even though our measurements are in 3D, we map them and proceed to analyze them in 2D. The conjecture is that 3D capture of face surface data contains richer information, while the 2D capture of the face images would have sacrificed this additional information in the first place. Notice that previous recognizers of prototypical expressions operate entirely in 3D [18, 12, 6, 7, 13, 14]. However, these methods depend on a large number of facial points, and are therefore in a different category and would not allow a fair 2D-to-3D comparison. There are several benefits of capturing facial data in 3D and then process them in 2D: Proven 2D AU methods can be applied directly, the higher computational load of 3D is reduced, and we have opportunity to compare the performance of 2D and 3D modalities under the same set of algorithms. To have adequate mapping of 3D facial surface geometry on the 2D image domain, surface curvatures are used. They provide faithful information about the geometry of the surface in a compact form, and thus can be very useful in facial deformation analysis, as has already been shown in [10].

Before curvature estimation, the 3D face data is subjected to certain pre-processing steps. First, 3D facial surface is reconstructed in terms of piecewise planar surfaces of a wireframe structure from 3D coordinate data (Figure 1b). Since 3D sensing devices produce noisy data and curvature estimation is noise sensitive, several noise filtering steps are run to remove spikes, to smooth data, and to fill in the holes (Figure 1c). Then, curvatures are estimated on the wire-frame surface and mean curvature values are resampled in the image domain via orthogonal projection. The final curvature image is obtained by applying image extrapolation for the out of domain areas [15] (Figure 1d). The resolution is the same with of the 2D camera image data, i.e., 96×96 pixels.

4. Experimental Results and Discussions

In order to evaluate 3D modality for facial expression analysis, we prepared an extensive database (Bosphorus Database [9]) where 3D faces were acquired with a structured light system. This database contains 105 subjects acting a large repertoire of expressions, and show various head poses and occlusions (beard, moustache, glasses, etc.). There are up to 54 face images per subject. The facial expressions were instructed by the experimenter and the ground-truth FACS codes were obtained by a certified FACS coder. The acquisitions were done under good illumination conditions. The color images have 1600×1200 resolution and the number of points on 3D faces varies roughly between 30K and 50K. Some sample 3D faces are shown in Figure 2.

In addition to the Bosphorus database, Cohn-Kanade DFAT-504 [4] database was also employed to evaluate 2D performances. It is one of the most common AU coded facial expression databases. It contains digitized video clips at 640×480 pixels resolution. Up to 23 instructed facial expressions per subject were captured from 100 university

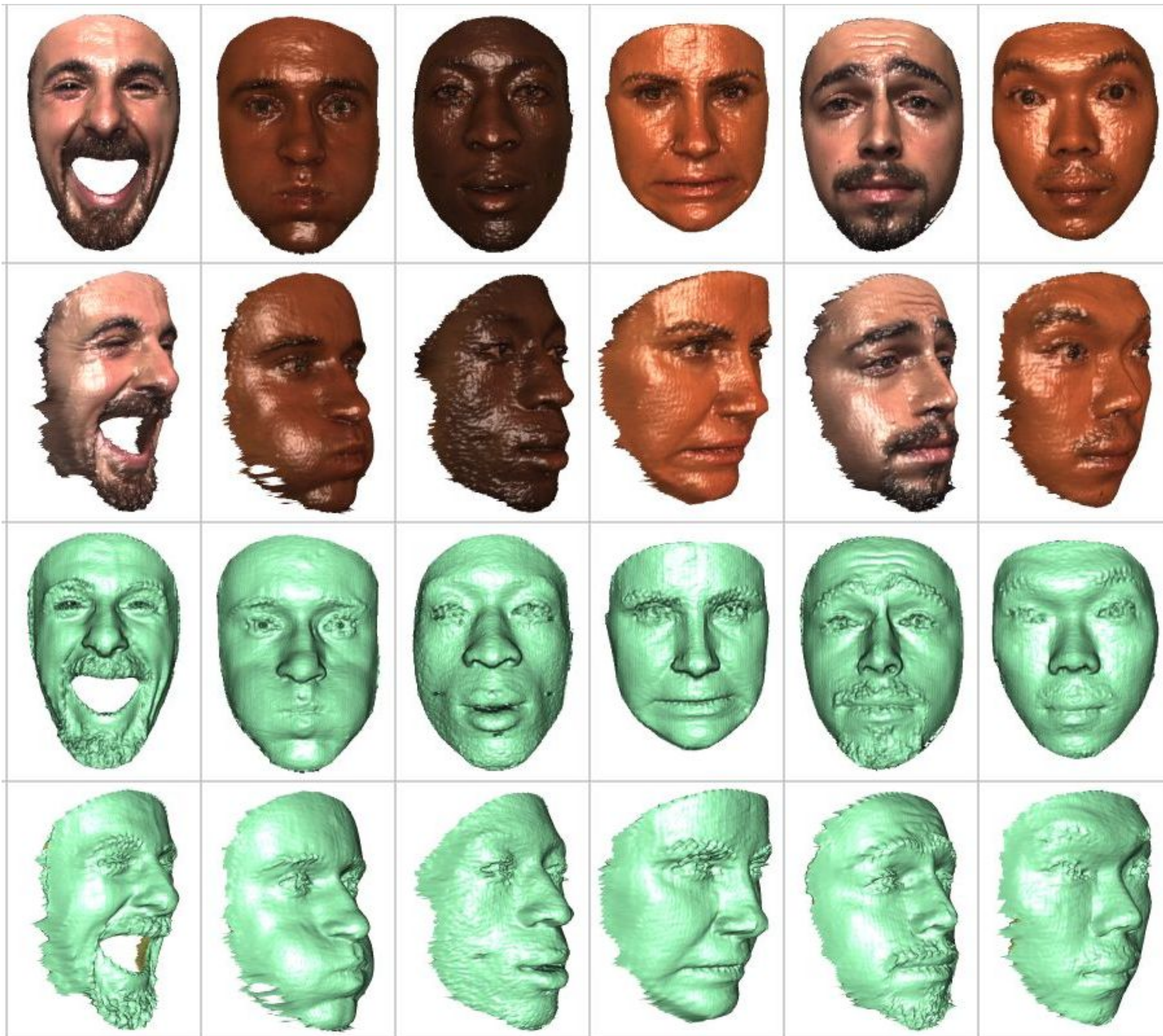


Figure 2: Sample 3D faces from Bosphorus database. The samples are shown with and without texture mapping by artificial lighting.

student. Each clip starts with neutral and ends with expression apex frames. 972 images of these neutral and apex frames has been used to test 19 AUs from this database.

In the Bosphorus database, 25 AUs, split as seven lower AUs and 18 upper facial AUs, are tested over 2902 samples in the experiments. Some examples of these lower and upper facial AUs are shown in Figure 3 and Figure 4 respectively. Each AU is treated separately; in other words, we develop 25 detectors, one for each AU. Any face image involving the target AU, alone or in combination with other AUs, is treated as a positive sample of that AU class, while all other images that do not involve the target AU are accepted as negative samples. All the experiments are performed using 10-fold subject cross-validation so that samples from any training subject are never used in testing. However, this dataset partitioning is not trivial since AUs are not distributed evenly among the subjects. We solve this problem by creating different subject partitions for each AU so that each fold becomes balanced with respect to the positive samples.

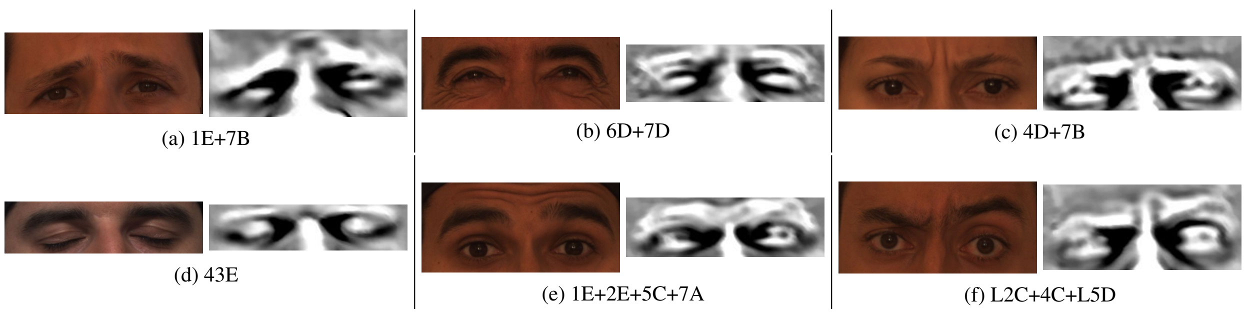
The performance of the detectors is given in terms of cross validated Receiver Operating Characteristics (ROC) curves. ROC curves show hit rate versus false alarm rate under varying thresholds. To have a single figure of merit that summarizes a ROC curve, Area under the Curve (AuC) measure is used, since it is equivalent to the theoretical maximum achievable correct rate of a binary classification problem. In the following subsections, we compare the pros and cons of the 3D and 2D modalities.

(a) 23C (b) 20C+23C+25C (c) 12B+14C+23C

(d) 14C+23C (e) 18D+24D (f) 9D+17C

(g) V10B+16D+25C (h) 12E+16D+25E (i) 22D+25D+27C

Figure 3: Camera and curvature images are shown for some lower facial action units. The FACS codes are also given.



(a) 1E+7B (b) 6D+7D (c) 4D+7B

(d) 43E (e) 1E+2E+5C+7A (f) L2C+4C+L5D

Figure 4: Camera and curvature images are shown for some upper facial action units. The FACS codes are also given.

4.1. Conjecture: 3D Expression Data is Better than 2D Data

Every AU in the Bosphorus dataset is annotated with its intensity level. This enables us to examine the detection performance as a function of the intensity. The five intensity levels of FACS range from slightest level (A to strongest level E). In our experiments we excluded the AU samples with intensity level A, since even the expert annotators are not very sure of their scores. The low B level intensity AUs are tested separately to see the effect of intensity. Thus, we developed the AU detectors using AUs with intensity levels C,D, and E for training and AUs with intensity levels B, C, D and E for testing.

The best results were obtained with linear SVM, outperforming AdaBoost and Gaussian Bayes classifiers. The AuC performance was 93.7% for 2D camera data of Cohn-Kanade dataset, 93.5% for 2D Bosphorus dataset and it was 2% points higher with 3D data, that is, 95.4%. These averages have been calculated by weighted summations of individual AU scores according to number of positive samples in that AU set. However, evaluations with only the average results does not show all the interesting information. To this effect, we show the detection rate of each AU, in Figure 5 to compare the 2D datasets, and in Figure 6 to compare 2D and 3D modalities. The AU bars are ordered according to differences in the performance. The 2D results of the two databases seem to be very close, but interestingly differences for some of the AUs are very big. One reason of this may be difference in the number of available positive samples. We observe that the advantage of 2D and 3D modalities differ from AU to AU. To interpret these figures, consider the first bar that belongs to AU23 and of which there are 63 realizations in the dataset. The darker part of

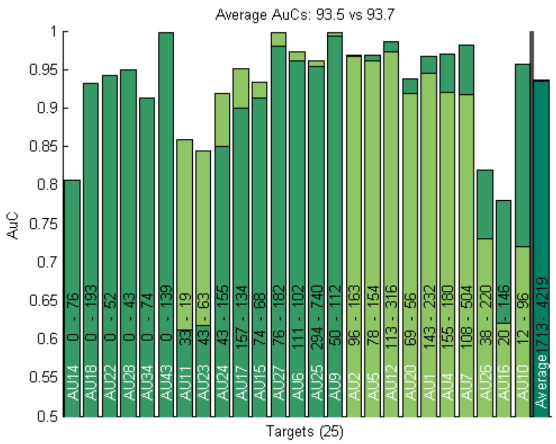


Figure 5: The performance comparison between 2D datasets of Cohn-Kanade and Bosphorus databases. SVM classifiers are used and AuC values are reported in bars (light color: Cohn-Kanade, dark color: Bosphorus). Number of total positive samples is written inside the AU bars, at the bottom for Cohn-Kanade, and at the top for Bosphorus dataset.

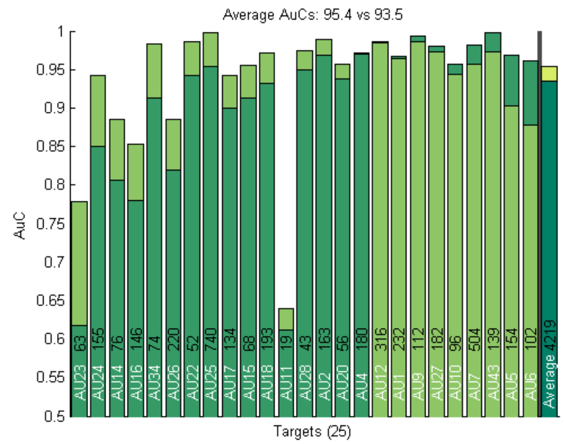


Figure 6: The performance comparison between 3D and 2D data modalities. SVM classifiers are used and AuC values are reported in bars (light color: 3D data, dark color: 2D data). Number of total positive samples is written inside the AU bars.

the bar indicates that 2D data achieves only 62% correct detection while 3D data achieves 87%; conversely, consider the last AU bar (AU16 with 102 instances in the dataset). In this case 3D data achieves 86% while 2D is better at 96% correct detection rate. We can conclude that in general 3D data considerably improves the detection of lower facial AUs. For example, the detection of AU23 (Lip Tightener), which was shown to be [1] one of the most difficult AUs, improves considerably, with a 16% gain. Several instantiations of this AU are shown in Figures 3a, 3b, 3c and 3d. Improvements on AUs 24, 14, 16, 24 and 26 are also remarkable.

3D is not necessarily always more advantageous. We can see that some performance degradations occur on upper facial AUs with 3D data. This may be explained by two factors. First, 2D eye texture and especially appearance of pupils are very informative for AU scoring. Second, eye region can be quite noisy with structured light based 3D acquisition due to eyelashes and glitters. This may hide necessary surface detail for detection.

Finally, AU11 (Nasolabial Furrow Deepener) deserves some explanation as this is a case under both modalities. The reason may be that too few positive samples are available in the dataset.

4.2. Performance with Action Units at Low Intensity

In real life, facial expressions can often occur at low intensities. Naturally, this makes the already challenging AU detection problem even more difficult as differences among AUs or between AU and neutral state become more subtle. Bartlett *et al*. [1] found 21% drop (from 92% to 71%) on average AuC values when their detectors are tested on a spontaneous expression database instead of posed expressions.

In order to see detection capabilities of lower intensity actions, we use only the B level samples as positives. The same subject cross validation partitions of the experiments in Section 4.1 are used, training sets are not modified, and also the negative test samples are the same. Hence, the only difference in the setup is the positive test samples. Figure 7 shows the results of 3D and 2D data with linear SVMs. We see a severe performance drop on average AuCs in both modalities: from 93.5% to 80.1% for 2D, and from 95.4% to 83.7% for 3D. Only AU9 (Nose Wrinkler) keeps its high detection rates under either modality. The improvement with 3D data is higher in case of lower intensity actions, from 1.9% to 3.6%. When we compare the differences between 3D and 2D on AU basis, we can come to conclusions similar as those given in Section 4.1. However, for AUs 25 (Lips Part), 22 (Lip Funneler) and 4 (Brow Lowerer) the detection rates drop considerably with 2D data while 3D detectors are still performing well. On the other hand, degradation in 3D detector of AU43 (Eye Closure) is much more than its 2D opponent, which may be related to eye texture and 3D acquisition noise as mentioned in Section 4.1.

4.3. 2D versus 3D Under Different Classifiers

We compare 2D and 3D modalities under different classifiers. The average AuC results are given in Table 1, and Table 2 lists these results for only B intensity level AUs.

Classifier	Cohn-Kanade - 2D	Bosphorus - 2D	Bosphorus - 3D	2D-3D Fusion
AdaBoost	92.1	92.2	94.8	95.5
Linear SVM	93.7	93.5	95.4	**97.1**
G-1 (Quadratic)	90.9	90.0	93.1	94.9
G-2 (Simple Quadratic)	76.4	63.3	85.4	92.1
G-3 (Naïve Bayes)	93.5	91.4	95.3	96.7
G-4 (Nearest Mean)	87.8	82.6	92.0	96.7

Table 1: Average AuC values of AU detectors for 2D and 3D data under different classifiers. All of the classifiers use 200 Gabor features that are selected by AdaBoost, except the fusion where concatenated 400 length feature vector is employed.

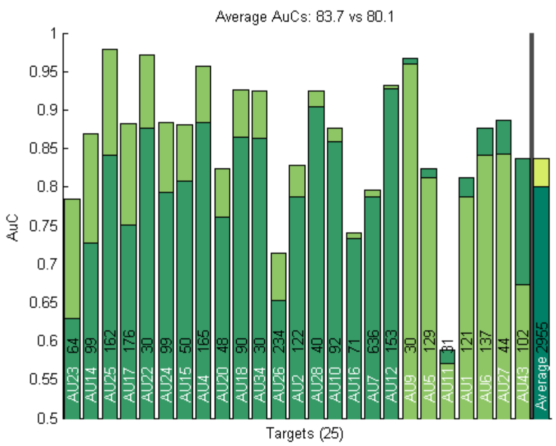


Figure 7: Low intensity detection (B level) performance comparison between 3D and 2D data modalities. SVM classifiers are used and AuC values are reported in bars (light color: 3D data, dark color: 2D data). Number of total positive samples is written inside the AU bars.

Two immediate observations are that 3D modality is performing better than 2D under any classifier, and that among the classifiers linear SVM is the winner. However, we observed that for some of the AUs which have less samples, Naïve Bayes performs better than linear SVMs over the 2D datasets. Another interesting observation is that while discriminative classification is apparently better for 2D data, for 3D data generative classification with Naïve Bayes is better than AdaBoost and almost the same with the SVM. Naïve Bayes assumes Gaussian distribution of the samples and linear separation. However, SVMs can potentially handle more complex and nonlinear discrimination boundaries by performing linear separation on higher dimensions. A tentative interpretation of this outcome is that facial features of 2D data have more complex distributions and not as well linearly separable as the 3D case. There is inherently nonlinearity in the 2D luminance images since they are non-linear measurements of facial surface, thus feature distributions of different classes of facial surface deforma-

Classifier	2D	3D	Fusion
AdaBoost	78.6	81.7	82.6
Linear SVM	80.1	83.7	**85.6**
G-1 (Quadratic)	76.9	79.5	81.8
G-2 (Simple Quadratic)	56.7	69.7	78.0
G-3 (Naïve Bayes)	79.2	81.9	84.8
G-4 (Nearest Mean)	68.7	77.0	84.8

Table 2: Detection performances on lower intensity B level AUs by average AuC values of AU detectors for 2D and 3D data under different classifiers. All of the classifiers use 200 Gabor features that are selected by AdaBoost, except the fusion where concatenated 400 length feature vector is employed.

tions may in turn not be as well linearly separable as the 3D case. We also observe that use of global variance estimates, as in the case of G-4 (Nearest Mean) and G-2 (Simple Quadratic), is not severely diminishing the performance with 3D data compared to 2D. This is also an indication for even simpler distribution of 3D data features.

4.4. Fusion of 2D and 3D Modalities

Finally, we evaluate the fusion of the two modalities, since 2D camera images and 3D data may contain complementary information that can be an aid in AU detection. It is realized by directly concatenating the 2D and 3D data features which are selected by AdaBoost. From Tables 1 and 2, we see very remarkable performance rises when the features of 2D and 3D modalities are combined. SVM achieves the best performance score of 97.1%, and for low intensity AUs, it is 85.6%. The results of Bayes classifiers are also high, and interestingly, even the worst performing Simple Quadratic classifier demonstrates high rates with fusion.

5. Conclusion

We investigated the use of 3D data for the facial action unit detection in lieu of conventional 2D luminance camera images. The two modalities were extensively compared using ROC analysis of 25 selected AUs. For fairness of comparison, the 3D data is converted to 2D images of surface curvatures. Once the two competitor modalities are in the form of 2D images (intensity versus curvature), we do Gabor analysis and select equal number of features via AdaBoost.

The results show that 3D data offers significant advantages in AU detection. In general, lower face AU detections benefit more from 3D as compared to 2D. Especially the detection rate of difficult AU23 improves considerably. 3D also proves its value for low intensity expressions. In the case of the next to the lowest intensity level (level B), while many AUs degrade for 2D, 3D data can maintain high performance for lower face AUs 25 and 22, and for upper face AU 4. Nevertheless, there are some upper face AUs where 2D outperforms 3D. We may explain the lower performance of 3D for upper face AUs with the fact that 3D sensing noise is excessive in the eye region, and misses the eye texture information. Another conclusion is that for 3D modality both generative and discriminative classifications have almost identical performances; in fact Naïve Bayes and linear SVM have scores of 95.3% and 95.4% respectively. Since neither 3D nor 2D is uniformly better, it is natural to think of their complementary roles. The feature fusion of the two modalities proves advantageous as the average AU recognition performance rises from 95.4% to 97.1%.

In a future study, we plan to evaluate other representations of 3D surfaces. For instance depth or Gaussian curvatures can be used as alternative representations of 3D geometry. Also, instead of 2D projection, parameterization techniques that establish bijective mapping can be employed in order to prevent information loss due to mapping from 3D to 2D domain.

A relevant future work is to observe the performance of 3D on spontaneously occurring AUs. However, this research has to await a 3D spontaneous expression database. Presently most of the existing 3D sensing devices use active sensors that project light on the subjects and capture only static data. However, recent progress in 3D acquisition, such as use of invisible spectrum and video acquisitions, may permit preparation of spontaneous databases for future research studies.

6. Acknowledgements

This work is supported by Bogazici University BAP 09HA202D and TUBITAK 107E001 grants. The numerical calculations reported in this paper were performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TR-Grid e-Infrastructure).

References

[1] M. S. Bartlett, G. Littlewort, M. G. Frank, C. Lainscsek, I. Fasel, and J. R. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, 1(6):22–35, 2006. 1, 2, 6

[2] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press, Palo Alto, 1978. 1, 2

[3] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2):124–129, 1971. 2

[4] T. Kanade, J. Cohn, and Y.-L. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, pages 46 – 53, March 2000. 3

[5] S. Lucey, I. Matthews, C. Hu, Z. Ambadar, F. de la Torre, and J. Cohn. Aam derived face representations for robust facial action recognition. In *FG06*, 2006. 1

[6] I. Mpiperis, S. Malasiotis, V. Petridis, and M. G. Strintzis. 3d facial expression recognition using swarm intelligence. In *ICASSP 2008*, Las Vegas, Nevada, USA, 2008. 2, 3

[7] I. Mpiperis, S. Malassiotis, and M. Strintzis. Bilinear elastically deformable models with application to 3d face and facial expression recognition. In *FG08*, 2008. 2, 3

[8] M. Pantic and I. Patras. Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. *Systems, Man, and Cybernetics, Part B, IEEE Trans. on*, 36(2):433–449, 2006. 1

[9] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun. Bosphorus database for 3d face analysis. In *BIOID 2008*, Denmark, 2008. 2, 3

[10] A. Savran and B. Sankur. Automatic detection of facial actions from 3d data. In *ICCV09: Workshop on Human Computer Interaction*, Kyoto, Japan, 2009. 2, 3

[11] M. A. Sayette, J. F. Cohn, J. M. Wertz, M. A. Perrott, and D. J. Parrott. A psychometric evaluation of the facial action coding system for assessing spontaneous expression. *Journal of Nonverbal Behavior*, 25(3):167–185, 09 2001. 1, 2

[12] H. Soyel and H. Demirel. Facial expression recognition using 3d facial feature distances. In *ICIAR07*, 2007. 2, 3

[13] H. Tang and T. Huang. 3d facial expression recognition based on automatically selected features. In *CVPR08: Workshop 3DFace08*, 2008. 2, 3

[14] H. Tang and T. Huang. 3d facial expression recognition based on properties of line segments connecting facial feature points. In *FG08*, 2008. 2, 3

[15] A. Telea. An image inpainting technique based on the fast marching method. *Graphics Tools*, 9(1):25–36, 2004. 3

[16] Y.-I. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, 23(2):97–115, Feb 2001. 1

[17] F. Tsalakanidou and S. Malassiotis. Real-time 2d+3d facial action and expression recognition. *Pattern Recognition*, 43(5):1763 – 1775, 2010. 2

[18] J. Wang, L. Yin, X. Wei, and Y. Sun. 3d facial expression recognition based on primitive surface feature distribution. In *CVPR '06*, Washington, DC, USA, 2006. 2, 3