Facial Action Unit and Emotion Recognition with Head Pose Variations

Chadi Trad[1], Hazem Hajj[1], Wassim El-Hajj[2], and Fatima Al-Jamil[3]

[1] American University of Beirut, Faculty of Engineering and Architecture, Department of Electrical and Computer Engineering
[2] American University of Beirut, Faculty of Arts And Sciences, Department of Computer Science
[3] American University of Beirut, Faculty of Arts and Sciences, Department of Psychology
{cht02,hh63,we07,fa25}@aub.edu.lb
http://www.aub.edu.lb/

Abstract. Facial expression recognition has been an active research topic for many years, with Facial Action Coding Systems (FACS) being among the widely used methods. FACS is a well-established scheme in psychology to annotate facial muscle contractions and relaxations, also called Action Units (AUs). Previous work on FACS-based methods focused on frontal or near-frontal head poses. In this work, we propose a method to recognize expressions in side head poses. This method builds one classifier for each possible group of occlusions. Facial expression recognition of a side facial pose is then based on a boosting approach of the different classifiers. The method is first tested with frontal and near-frontal head poses, and the results are shown to be comparable to state of the art work for AU and emotion detection. The method is then tested with a small training set for various orientations and AUs, and shown to be accurate.

Keywords: Facial Action Coding System (FACS), facial expressions, head pose, non-frontal, Action Unit (AU).

1 Introduction

Traditional Human Computer Interface (HCI) designs have focused on interface devices that convey explicit messages from the user, such as touch screens and keyboards, while the affective state has been ignored. However, recent research has demonstrated the importance of affective state in human-human communication [1]. Therefore, recognizing the user's affect can make HCI more natural.

In facial expression detection, the research areas are generally divided into two main streams: facial expression measurement and facial muscle action detection, which fall under the categories known as message-judgment and sign-judgment approaches, respectively [2]. While the aim of the former is to detect the affect underlying the facial expression, the latter aims to purely describe the state of facial components such as their movements or shapes, leaving affect judgment to a higher level process. Facial Action Coding System (FACS) is considered as

S. Zhou, S. Zhang, and G. Karypis (Eds.): ADMA 2012, LNAI 7713, pp. 383–394, 2012.

a sign-judgment approach. FACS, which was introduced by Ekman et al. [3], is the most used coding scheme that describes the muscular activity of a face. Indeed, FACS describes visually discernible facial movements in terms of action units (AUs). Ekman and Friesen identified 44 AUs, which were associated with the contraction of facial muscles. They also provided rules for recognition of the onset (start), apex (peak) and offset (end) of the AUs. After detecting the AUs in the face, the universal emotions of Ekman, such as anger, disgust, and happiness, can be recognized by using the Emotional FACS rules (EMFACS). In addition, more affective states can be also recognized using FACS Affect Interpretation Database (FACSAID). Consequently, using this representation simplifies the affect recognition problem to the muscular structure of the face, thus, reducing thousands of facial expressions to a combination of few AUs.

Generally, facial expression recognition consists of two phases: feature extraction and classification stage. A large number of methods have been developed for each phase. For instance, feature extraction is divided into two types of approaches: holistic techniques (e.g. [4]) and local techniques (e.g. [5] and [6]). While holistic approaches use the entire face as input, the local approaches only consider specific features to be used such as geometrical features and appearance features or a combination of both. On the other hand, the classification techniques can be divided in two groups: spatial approaches [4] and spatio-temporal approaches [5]. While spatial approaches consist of analyzing the features on a frame by frame basis, spatio-temporal methods consider the evolution of the features in time. Finally, we note that improving feature extraction and classification techniques can improve the accuracy of automatic AU recognition systems, independently of the application. Furthermore, systems that can consider realistic situations, such as various head poses and occlusions, are more applicable to real life applications, since such difficulties occur in naturalistic settings. Most of the current methods proposed are robust to small head pose variations, such as the work of [5]. Few researches have worked on 3D databases, such as [7] and [8], but have focused on identifying the universal emotions rather than the FACS AUs. In this study, we propose a system that can recognize FACS AUs for various poses. Rather than training one classifier per AU for the whole face, we train multiple classifiers for each part of the face. The final decision combines the classification results while considering which parts of the face were occluded. The rest of this paper is organized as follows. In Section 2, we present related work and background on our proposed system. In Section 3, we introduce our proposed method. Experimental results are shown in Section 4 before drawing conclusions in Section 5.

2 Related Work and Preliminary Concepts

In this section, we first present related work for the particular area of facial expression recognition for various poses as well as geometrical-feature-based methods. Then, a background on the methods used is provided. In AU feature extraction, two types of local features are generally used: geometrical and appearance based features, or a combination of both. In this work, we focus our experiments on geometrical features to demonstrate the overall performance of our system. However, the same proposed approach can be used for a combination of geometrical and appearance based method.

2.1 Related Work

Virtually most of the AU methods reported so far have been based on near frontal views data [5] or on data with moderate head pose variation [9]. For instance, the work of Pantic et al. [5] investigated facial AU recognition from near-frontal views using geometrical features, and modeled the temporal phases of AUs. One of the weaknesses reported is that significant out-of-plane rotations affected the recognition accuracy. Other work such as the work of Tyan et al. [6] used a combination of geometrical and appearance based features. This method was reported to be robust for moderate face rotations, but no direct measure for the accuracy/angle dependency was given.

To extract facial expressions in less-constrained environments, such as different head poses, Pantic and Patras [10] investigated facial AU recognition from profile views. Also, several experiments were done on the BU-3DEF [11] database, which is a facial expression database that consists of 3D shapes and 2D facial textures from 100 subjects. One of these works is the work of Hu et al. in [7] where they studied facial expression recognition for different head orientations, i.e. yaw and pitch angles, by extracting appearance based features such as Histogram of Oriented Gradients (HoG), Local Binary Patterns (LBP) and Scale Invariant Feature Transform (SIFT). These descriptors are well known in the area of computer vision for object and texture classification. In [8], Hu et al. use the displacement of manually selected feature points from the face for classification. However, all of these methods were tested on detecting the basic emotions of Ekman and not on detecting FACS AUs.

In the work of Valstar et al. [5], the authors extracted the facial points using a tracking scheme based on particle filtering using factorized likelihoods (PFFL). Affine transformation was then performed on the obtained coordinates to reverse the effect of scaling and small head orientations. Geometrical features as well as temporal features were extracted from the image sequences. Finally, a combination of Gentle-Boost and Support Vector Machine (SVM) was used in the classification stage. The system was further extended to detect the temporal activation model (neutral, onset, apex and offset). The main disadvantages of their system in detecting AUs for pose variations can be summarized by the following: The affine transformation cannot model out-of-plane rotations assigned with the head pose, and (2) PFFL cannot handle facial point occlusions associated with head pose variation.

2.2 Preliminary Concepts

In this sub section, we describe the general parts in a geometrical based AU recognition system and introduce the methods used in our system. First, a facial tracker is employed to detect and track the facial points. One of the most used models for facial tracking is the Deformable Model Fitting (DMF). DMF is a classic problem formulation in which the shape of object deformations is modeled using the Point Distribution Model (PDM) founded by Taylor [12]. In this model, the facial points' positions are calculated using the following equation:

$$\mathbf{x}_i = s\mathbf{R}(\bar{\mathbf{x}}_i + \mathbf{\Phi}_i\mathbf{q}) + t \ , \tag{1}$$

where $\mathbf{x}_i$ denotes the location of the i^{th} landmark, s denotes a scale, $\mathbf{R}$ a rotation matrix, $\bar{\mathbf{x}}_i$ the mean location of the i^{th} landmark, q a set of non-rigid parameters, and Φ a submatrix of basis variations. The aim is to determine the landmarks positions x_i, by determining the set of shape parameters (shape, rotation, translation and non-rigid parameters). Particularly, the Regularized Deformable Model Fitting (RDMF) tracker [13] follows the DMF model. RDMF uses a logistic regressor function to determine the likelihood of a facial point position, given an input image. The values of the landmark positions are determined by minimizing the misalignment error according to the PDM model as well as maximizing the new position likelihood. The search space of an optimal solution is minimized using hill climbing methods. The advantage of this tracker is that it is robust to multiple occlusions since it leverages the relationship among the facial points in the PDM model. After extracting features from the obtained facial points' positions, a machine learning algorithm such as SVM can be used to classify an AU. However, if the feature dimension is greater than the training data, overfitting to the training data is rather probable. Many feature reduction techniques can be used at this stage. Gentle-Boost combines a weighted vote of weak classifiers in the final classification. In the next section, we propose how to combine RDMF with Gentle-Boost to detect the AUs for various orientations.

3 Proposed Method

In this section, we describe our proposed method illustrated in Fig. 1. The first step is to detect and track a set of facial point coordinates using the RDMF tracker proposed in [13]. These coordinates are then separated into two groups: left-face points and right-face points. The features extracted are the distances' variation among the points. Finally, we describe the model for detecting the activation status of each AU. In the following subsections, we first describe how the features are extracted for each face region, and then we explain the AU classification scheme. Finally, we describe a classification system for emotion detection.

3.1 Feature Extraction from Left and Right Face Regions

Each image sequence is first processed using the RDMF tracker [13] obtain facial points coordinates across all frames. We note the coordinates of these points as:

$$X = ((x_1, y_1), (x_2, y_2), ..., (x_n, y_n)) \ . \tag{2}$$

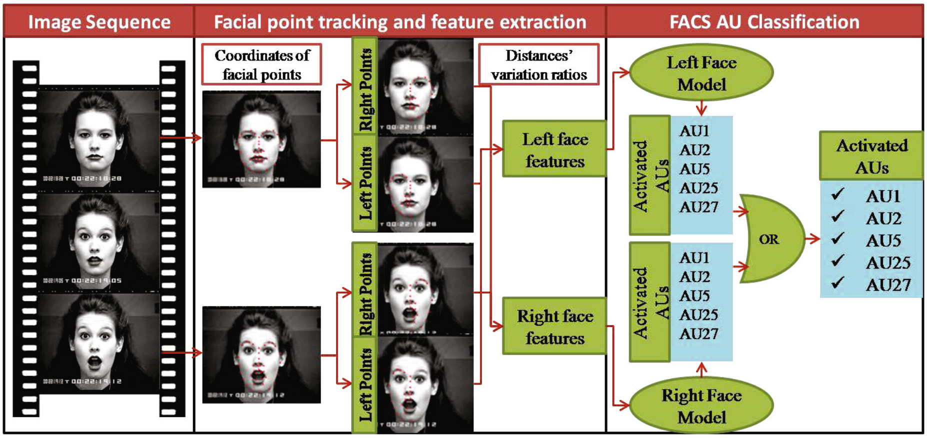

Fig. 1. AU detection algorithm

In this work, the number of facial points is 66, which is based on the RDMF implementation. We note at this point that not all of these points can be extracted when an area from the face is occluded, such as the right or the left facial area. After detecting the facial points, a set of features are then extracted. Euclidean distances d_{ij} among the points are calculated, such that: $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. Then, the set of features are extracted from the specified facial area. The features are defined by the ratios:

$$R_{ij} = \frac{d_{ij}}{d^r_{ij}} \tag{3}$$

where d^r_{ij} is the distance between the facial points i and j in a reference frame: a frame where the facial expression is neutral. The choice of these features is suitable when the head poses in the neutral frame and the tracked frame are relatively close. Rather than employing all the points in the AU classification, we propose to train multiple classifiers for each facial area (in this case left or right). It follows that only pair of points within one part of the face are used as features in each classifier. For instance, consider a facial tracker that can track three disjoint sets of facial points: **A**, **B** and **C**, and assume that either one of the set of points **A** or **C** can be occluded at once (for example the left or right facial area). Rather than training a single classifier $\mathbf{M_{ABC}}$ on all the landmarks that belong to **A**, **B** or **C**, we propose to train two classifiers: $\mathbf{M_{AB}}$ and $\mathbf{M_{BC}}$, that is for all possible combinations of landmarks being present/absent. For instance, in the testing phase, in the case where **A** (respectively **C**) is occluded, $\mathbf{M_{BC}}$ classifier should be used (respectively $\mathbf{M_{AB}}$). The occlusion status of the facial area can be directly extracted from the RDMF tracker by checking the point coordinates. On the other hand, if no landmarks from **A** or **B** are occluded, the decision should be weighted between $\mathbf{M_{BC}}$ and $\mathbf{M_{AC}}$. The same procedure applies for the training phase, where the classifiers can be trained only when their corresponding sets are not occluded. In our implementation, we consider three sets of landmarks: left-face-only set **L**, right-face-only set **R** and common points **J**. In the case where no area is occluded, the final decision is based on a logical OR between $\mathbf{M_{LJ}}$ and $\mathbf{M_{RJ}}$. In the remaining part of the paper, the term "left face" (respectively "right face") will refer to the points in **L** and **J** (respectively to the points in **R** and **J**). It is worth noting that the difference between this method and conventional ones is that multiple classifiers with various features are being used for each face region rather than using one classifier for the whole face.

3.2 Action Unit Detection Model

The training algorithm for each AU classifier is illustrated in Fig. 1 and Fig. 2. In the training phase, the neutral and apex frames are extracted from each video. The features from the left and right areas of the face are collected separately. One classifier is trained to classify the activation state of each AU (activated or not) and for each area of the face, using the features collected. The activation state of each AU should be available in the database or manually annotated by a FACS coder. We employ Gentle-Boost algorithm to avoid data overfitting on one hand, since the number of features is higher than the number of training data, and to make our work more comparable with other works in the literature. In the testing phase, if a facial area is occluded, the classifier of the other area will be used for classification. In the case where no facial area is occluded, the activation state of the AU is calculated by performing a logical OR on both left and right classifications. In fact, the FACS manual states that if an AU is activated in one part of the face, e.g. left eye brow raiser, the AU is annotated to be activated.

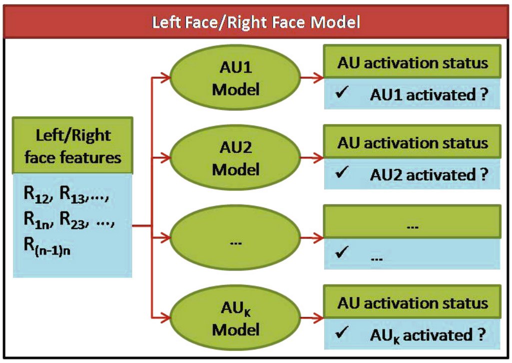

Fig. 2. Left/Right face model

3.3 Emotion Detection Model

In order to detect emotions, the same features extracted from the face are used to detect the various emotions expressed by the subjects: anger, disgust, fear, happiness, sadness, surprise or neutral. A number of classifier are trained using the Gentle-Boost algorithm to differentiate between each pair from the set of emotion (one-versus-one approach). The detected emotion in the left and right parts of the face cannot be combined together in the same way the detected AUs were combined in the previous section (using the logical OR). In the experiments section, we show the results of each region separately.

4 Experiments and Evaluation

In our system, we employed the author's implementation of the RDMF facial tracker in our system [13]. The code executes in real time and its output ranges from 20 - 30 fps based on the processor and the compiler used. In order to evaluate our method, we perform three experiments. In the first one, we test our system on the Cohn-Kanade (CK) database [12] which contains 480 gray scale videos that were made public. The head orientation of subjects in the recorded videos is near-frontal. This database was collected for the purpose of facial expression recognition and this is currently the most used database in this research area. We study our system on this database in order to validate our results by comparing them with a state-of-the-art geometrical approach for small head orientations. In this comparison, we use the results obtained in the frame-based experiments by Valstar et al. in [5]. In the second experiment, the Emotion detection system is tested against the one proposed also in [5]. Finally, we validate our system in the third experiment for various poses of the head, ranging from 0 degree to 90 degrees. For this case, multiple videos were recorded featuring different orientations.

4.1 Benchmarking for Action Unit Detection

In the CK database, for each sequence of images, the facial landmarks were detected using the RDMF tracker. The coordinates of the 66 facial landmarks are tracked through each image sequence in the database. The left face area consists of 37 points while the right one contains 38. The numbers were determined experimentally and depend on the training of the facial tracker.

For each facial area, the ratios of the distances, described in section 3, were extracted from the neutral face and the apex frames. In the CK database, rather than manually labeling the apex frames for each AU in each video, which is time consuming, we considered the neutral and the apex frames to be the first and the last frames of the image sequence, respectively. We note that this assumption is fair for most AUs in this database, since most image sequences were recorded till the beginning of the apex stage for all AUs. However, we are aware of this assumption's limitation since it is not always valid, especially for certain AUs like AU45 (blink). As per features and class labels, the left face features and the right face features were extracted for each image as specified in the previous section. In summary, the size of the feature sets are presented in Table 1. Finally, we mention that we trained each of the Gentle-Boost classifiers for 10 rounds.

Table 1. Size of the facial landmark sets and their corresponding features

Size of landmark sets			Size of feature vector		
Full-Face	Left	Right	Full-Face	Left	Right
66	37	38	2145	666	703

We conduct our experiment using the leave-one-subject-out strategy. In each fold, one subject is left out of the database, and all classifiers are trained on the remaining subjects and then tested on the subject that was left out. Binary confusion matrices are then summed together for all the experiments, i.e. for each subject in the database. Table 2 shows a comparison between the results obtained in our work, named Distance Ratio Classifier (DRC), and the one in the paper [5] by Valstar et al., illustrated as (TMP). In the third column, the number of positive examples for each AU is illustrated. All AUs that were previously studied, except for AU10, are also studied in this work. In our experiment, all 500 image sequences from the CK database were used, whereas the number of image sequences used in the TMP algorithm is 153. Four measures were calculated: accuracy, recall, precision and F1 measure. While the accuracy measure is a highly biased measure due to the unbalanced nature of the data, precision and recall are a better approximation of the data. The F1 measure combines the two latter measures by favoring them equally. The table is interpreted as follows. For each AU, compare the F1 column of the DRC and TMP algorithms. Precision and recall can be used for further investigation on the property of the classifier used. When needed, we refer to the accuracy of column. However, we note again that the latter measure is not very significant since the number of negative examples for each AU is much bigger than the ones with positive examples. Although, the results obtained in our method are highly optimistic, no direct conclusion on the superiority of our algorithm over the temporal based algorithm can be made since the selection of videos used is not the same. As can be seen, AU1 (inner brow raisers) and AU24 (lip pressor) show very close results with superior measures for the DRC method. Additionally, our method is also superior for other AUs such as AU2, AU4, AU5, AU7 and AU9. We believe that the reason for this improvement is behind the DMF model used by the Facial Tracker. On the other hand, other AUs such as AU6 (cheek raiser) and AU12 (lip corner puller) show that the method proposed is not accurate. In fact, it can be observed that the landmarks of the lip corners are not tracked effectively using the RDMF when performing AU12. Lastly, the poorest result achieved by DRC is for AU45 (blink). The lack of precision for this AU detector is mostly attributed to the preprocessing assumption that we made. In fact, most image sequences end after

Table 2. Comparison between the work in this paper (DRC) and the one in [5] (TMP). The number of videos used in the test is specified in the third column.

AU	Meth.	Videos	Acc.	Recall	Prec.	F1
1	DRC	144	0.910	0.809	0.864	0.835
	TMP	68	0.918	0.808	0.844	0.826
2	DRC	97	0.964	0.871	0.946	0.907
	TMP	50	0.939	0.791	0.879	0.833
4	DRC	156	0.896	0.755	0.864	0.806
	TMP	54	0.870	0.604	0.658	0.630
5	DRC	78	0.926	0.708	0.761	0.734
	TMP	37	0.904	0.566	0.629	0.596
6	DRC	111	0.870	0.713	0.694	0.703
	TMP	39	0.930	0.789	0.811	0.800
7	DRC	108	0.862	0.685	0.679	0.682
	TMP	31	0.870	0.268	0.315	0.290
9	DRC	50	0.972	0.864	0.826	0.844
	TMP	30	0.928	0.676	0.497	0.573
12	DRC	113	0.904	0.780	0.780	0.780
	TMP	42	0.930	0.827	0.844	0.836
15	DRC	81	0.910	0.609	0.661	0.634
	TMP	19	0.969	0.500	0.283	0.361
20	DRC	70	0.924	0.638	0.772	0.698
	TMP	34	0.908	0.466	0.582	0.517
24	DRC	43	0.928	0.421	0.533	0.471
	TMP	17	0.935	0.395	0.497	0.440
25	DRC	303	0.888	0.917	0.905	0.911
	TMP	19	0.851	0.717	0.782	0.748
26	DRC	39	0.926	0.175	0.636	0.275
	TMP	27	0.902	0.336	0.380	0.357
27	DRC	77	0.972	0.919	0.895	0.907
	TMP	30	0.964	0.836	0.873	0.854
45	DRC	19	0.954	0.091	0.400	0.148
	TMP	23	0.943	0.584	0.408	0.480
DRC Avg.			0.920	0.664	0.748	0.689
TMP Avg.			0.917	0.611	0.619	0.609

the offset of this AU has occurred, i.e. the final frame does not generally contain the apex for AU45. Any accurate result for this AU is due to the mere correlation between AUs in the database.

4.2 Benchmarking for Emotion Detection

In this section, the same features extracted previously are used in the emotion detection. We compare our method to the one proposed by Valstar et al. in [5] on the CK database. Note that the annotation for the emotions is provided in the database. The confusion matrix obtained in [5] is illustrated in Table 3(a).

The results obtained using our method are illustrated in Table 3(b). We test our system by using the points from left face only. The classification accuracy in our method is very comparable to the one in [5]. The classification rate for the emotion anger and surprise is much better in our method. On the other hand, the sadness classification rate is lower in our case. We note that the database subsets used are not the same in our experiment and the one in [5]. Thus, we don't elaborate more on the comparison, and simply state that the results obtained in our method are comparable to the state of the art approach in [5].

Table 3. (a) Confusion matrix for emotion classification using the method in [5]. (b) Confusion matrix for emotion classification using the features from the left face.

	An.	Di.	Fe.	H.	Sad.	Sur.	Rate
Ang.	2	3	2	0	9	1	0.118
Disg.	1	19	1	1	4	1	0.704
Fear	1	4	15	5	2	1	0.536
Hap.	1	0	3	33	0	1	0.868
Sad.	4	2	1	0	16	1	0.667
Sur.	0	1	1	1	0	34	0.919

	An.	Di.	Fe.	H.	Sad.	Sur.	N.	Rate
Ang.	19	3	0	2	4	0	1	0.655
Disg.	1	32	0	0	0	0	1	0.941
Fear	0	1	14	1	1	0	0	0.824
Hap.	0	0	1	60	0	0	0	0.984
Sad.	4	0	2	0	9	0	1	0.563
Sur.	0	0	8	0	0	63	0	0.887
Neu.	0	0	0	0	0	0	228	1

4.3 Various Pose Evaluation for Brow Raisers

In this section, we test our method on various orientations. For this purpose, a training set was created featuring two subjects in 45 videos in total. The sequences were recorded for three discrete yaw orientations (horizontal rotations), namely no yaw, moderate yaw and extreme yaw, approximated by: 0 degree, 45 degrees and 90 degrees angles to the camera imager. All videos were taken using a DMC-F3 Panasonic camera at a resolution of 1280x720 pixel^2, a rate of 30 fps and a distance of 2 meters from the subject.

The subject was asked to stand in a frontal pose, and then to rotate his head by a specific angle until facing a marker on the wall and to perform an AU or a combination of AUs. Afterwards, we manually annotated the neutral frames and apex frames of each sequence. We note that only the neutral frame preceding the onset phase is considered. A sample recorded set of images is shown in Fig. 3. The subjects were asked to perform AU1 and AU2 (brow raisers). As a matter of fact, it is essential to assess the validity of any FACS system for the most common AUs (AU1 and AU2 consists about 20% of the CK database).

In the testing phase, only the tracked points from the neutral and apex frame were extracted from the video. A previously trained DRC classifier set from the CK database was used on the data set. Table 4 illustrates the evaluation of our method for the three yaw intensities, and for the two AUs. We note that the tracker failed to track some videos for extreme face orientations. These videos are excluded from the final statistics. The second column shows the intensity of the head orientation. The numbers of positive and negative examples are illustrated

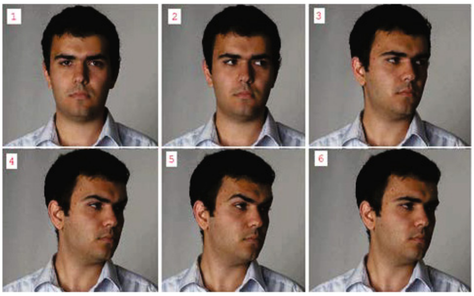

Fig. 3. Samples in an images sequence from the special database. The subject is first asked to rotate their face, then to perform the AU. The third image is annotated as the neutral face, and the fifth is annotated as the apex frame.

Table 4. Evaluating algorithm on the special database for 3 head orientation intensities and two AUs

AU	Int.	P	N	Acc.	Recall	Prec.	F1
1	1	4	10	0.929	1.000	0.800	0.889
	2	14	22	0.861	0.786	0.846	0.815
	3	9	11	0.950	0.889	1.000	0.941
2	1	4	10	0.643	0.500	0.400	0.444
	2	7	29	0.889	0.571	0.800	0.667
	3	5	15	0.850	0.600	0.750	0.667

in the third and the fourth column. The same measures from experiment 1 are used in this experiment. Not surprisingly, the F1 increased when the orientation intensity was stronger. In fact, this is consistent with the work in [10] which concluded that profile views are better than frontal views for AU detection.

5 Conclusion

Finally, we conclude our work with the following analysis. We have developed a working system that can detect AUs for various head poses without any prior training on these poses. Our method has comparable results with the state-of-the-art geometrical algorithm in [5] for near-frontal head orientation. Moreover, we realize that our model was able to generalize to a new database and that AUs were still detectable at various pose orientations. The main issue that we wish to address in our future work is to extract features that are more robust to pose variations. Lastly, using the tracker from [13] had its disadvantages, since some AUs couldn't be detected accurately, such as lip corner pullers. We wish to evaluate our work on other trackers that were previously used in the literature.

Acknowledgments. This work was funded by Intel-KACST Middle East Energy Efficiency Research (MER) and the American University of Beirut (AUB) University Research Board (URB). Special acknowledgment is due to Lama Nachman for her valuable inputs.

References

1. Russell, J., Dols, J.: The psychology of facial expression. Studies in emotion and social interaction. Cambridge University Press (1997)
2. Tian, Y.L., Kanade, T., Cohn, J.F.: Facial expression recognition. In: Li, S.Z., Jain, A.K. (eds.) Handbook of Face Recognition, 2nd edn., pp. 487–519. Springer (2011)
3. Ekman, P., Friesen, W.V.: The facial action coding system: A technique for the of facial movement. In: Consulting Psychologists (1978)
4. Lien, J., Kanade, T., Cohn, J., Li, C.: Detection, tracking, and classification of action units in facial expression. Robotics and Autonomous Systems 31(3), 131–146 (2000)
5. Valstar, M., Pantic, M.: Fully automatic recognition of the temporal phases of facial actions. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 42(1), 28–43 (2012)
6. Tian, Y., Kanade, T., Cohn, J.: Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2), 97–115 (2001)
7. Hu, Y., Zeng, Z., Yin, L., Wei, X., Zhou, X., Huang, T.S.: Multi-view facial expression recognition. In: FG, pp. 1–6 (2008)
8. Hu, Y., Zeng, Z., Yin, L., Wei, X., Tu, J., Huang, T.: A study of non-frontal-view facial expressions recognition. In: 19th International Conference on Pattern Recognition, ICPR 2008, pp. 1–4. IEEE (2008)
9. Tong, Y., Liao, W., Ji, Q.: Facial action unit recognition by exploiting their dynamic and semantic relationships. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(10), 1683–1699 (2007)
10. Pantic, M., Patras, I.: Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 36(2), 433–449 (2006)
11. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.: A 3d facial expression database for facial behavior research. In: 7th International Conference on Automatic Face and Gesture Recognition, FGR 2006, pp. 211–216. IEEE (2006)
12. Cootes, T., Taylor, C.: Active shape models–smart snakes. In: Proc. British Machine Vision Conference, vol. 266275. Citeseer (1992)
13. Saragih, J., Lucey, S., Cohn, J.: Deformable model fitting by regularized landmark mean-shift. International Journal of Computer Vision 91(2), 200–215 (2011)