

Automatically Detecting Action Units from Faces of Pain: Comparing Shape and Appearance Features

Patrick Lucey, Jeffrey Cohn, Simon Lucey
Robotics Institute, Carnegie Mellon University
Pittsburgh, PA, 15213 USA.

patlucey@andrew.cmu.edu, jcohn@cs.cmu.edu, slucey@cs.cmu.edu

Sridha Sridharan
SAIVT Laboratory, Queensland University of Technology
Brisbane, QLD, 4000, Australia.

s.sridharan@qut.edu.au

Kenneth M. Prkachin
Department of Psychology, University of Northern British Columbia
Prince George, BC, V2N4Z9, Canada

kmprk@unbc.ca

Abstract

Recent psychological research suggests that facial movements are a reliable measure of pain. Automatic detection of facial movements associated with pain would contribute to patient care but is technically challenging. Facial movements may be subtle and accompanied by abrupt changes in head orientation. Active appearance models (AAM) have proven robust to naturally occurring facial behavior, yet AAM-based efforts to automatically detect action units (AUs) are few. Using image data from patients with rotator-cuff injuries, we describe an AAM-based automatic system that decouples shape and appearance to detect AUs on a frame-by-frame basis. Most current approaches to AU detection use only appearance features. We explored the relative efficacy of shape and appearance for AU detection. Consistent with the experience of human observers, we found specific relationships between action units and types of facial features. Several AU (e.g. AU4, 12, and 43) were more discriminable by shape than by appearance, whilst the opposite pattern was found for others (e.g. AU6, 7 and 10). AU-specific feature sets may yield optimal results.

1. Introduction

Significant efforts have been made to identify reliable and valid facial movements associated with pain [5, 16, 17]. These movements include such actions as brow lowering, eye closure and cheek raising as defined by the Facial Action Coding System [6]. A major benefit of a system like FACS is that measures of pain can be detected at each time

step (i.e. each video frame), which is not possible using symptom rating scales. However, a limitation is that FACS must be performed offline, and manual observations are costly, which makes clinical use prohibitive. Observer- and patient rating scales are a less expensive alternative that can be used clinically, but they have other limitations [21], such as lacking the temporal precision of a facial-movement based system. A realtime system for detecting facial movements associated with pain would provide significant advantage in patient care and cost reduction.

In this paper we use the UNBC-McMaster Shoulder Pain Archive to develop an automatic system which we use to recognize individual action units (AUs) as defined by FACS on a frame-by-frame basis. Automatic AU detection is a first step toward an automated system of pain detection. In the UNBC-McMaster Archive, pain is real, not posed or feigned, the facial actions vary in duration and intensity, and facial actions are often accompanied by abrupt changes in non-rigid head motion. This makes AU detection especially challenging. To extract facial features and register face images to a canonical view, we use active appearance models (AAM) [4]. AAMs have proven to be robust to non-rigid head motion [13] and can decouple shape and appearance features. In this paper we analyze AU detection using shape and appearance both individually and in combination.

For human FACS coders, the relative importance of shape and appearance varies with type of AU. Brow lowering (AU 4) produces strong changes in shape and variable

changes in appearance. The mixed contribution of appearance features results from individual differences in facial furrows and wrinkles. Some people have a smooth brow at rest, while others have permanent facial furrows of different intensity and shape. Such individual differences can complicate the use of appearance features for AU detection. Cheek raising (AU 6), on the other hand, produces changes in shape that are easily confusable with closely related actions (AU 7 especially). Thus, the information value of shape or appearance for human FACS coders varies by action unit. An AAM and associated classifier may perform similarly. For an AAM, an additional consideration is that some AU coincide with the AAM mesh (e.g., AU 4), which allows them to be measured directly.

We test the hypothesis that the information value of shape and appearance features derived by an AAM varies specifically with type of AU. For example, we consider whether AU4 (brow lowering) might be more readily discriminated by using only a geometric representation (i.e. points on the mesh) rather than pixel information (appearance), which may be confusable with other facial actions. Conversely, the wrinkles that emanate from AU6 might be more discernible by using pixel (appearance) features rather than shape features, as these features might be more confusable with AU7.

1.1. Related Work

There have been many attempts to automatically recognize AUs, both individually and in combination. Bartlett and colleagues [2] developed a system that first detects the frontal face using the Viola-Jones detector [20], extracts visual features using Gabor filters, which are selected via AdaBoost and then train a support vector machine (SVM) to detect AU. A similar strategy has been employed by Littlewort et al. [8]. Tong et al. [19] also reported good performance with their system which uses a dynamic Bayesian network (DBN) to account for the temporal nature of the signal and the relationships among AUs. Pantic and Rothkrantz [14] used a rule-based method for AU recognition. Lucey, Ashraf, and Cohn [11] used active appearance model and SVM to detect action units in the brow region. Comprehensive reviews can be found in [18, 19].

Recent work has turned increasingly to the problem of spontaneous AU detection, in which facial actions occur in ecologically valid settings rather than being posed. AU detection in naturalistic settings is quite challenging because facial actions often coincide with out-of-plane head motion and consequent change in illumination. With few exceptions [3, 12, 11], previous literature in AU detection emphasizes appearance-based features and assumes frontal view with little head motion. Experience with FACS, however, suggests that shape features may carry additional information. Preliminary support for this hypothesis comes from Lucey and colleagues where they found that shape and

canonical appearance maximized AU detection in the brow region in subjects experiencing real-life pain [11]. Moreover, when facial actions are accompanied by moderate out-of-plane head motion, as in clinically observed pain, approaches that are robust to moderate head motion may prove beneficial.

In this paper we use AAMs to compare appearance and shape features in an attempt to discover which ones are more desirable for each individual AU. This comparison follows on from the work of [1] in which a comparison of shape and appearance features was performed at a sequence level to recognize pain. The rest of the paper is organized as follows. In the next section we describe the database that we use for our work. We then describe our automatic AU recognition system based on AAMs and SVMs, followed up by our results and concluding remarks.

2. UNBC-McMaster Shoulder Pain Expression Archive Database

The UNBC-McMaster Shoulder Pain Expression Archive database was used for this work. It contains video of the faces of adult subjects (129 subjects - 63 male, 66 female) with rotator cuff and other shoulder injuries. Subjects were recorded during movement of their affected and unaffected shoulder during active and passive conditions. In the active condition, subjects initiated shoulder rotation on their own. In the passive condition, a physiotherapist was responsible for the movement. In the experiments conducted in this paper, only the active condition was used. Within the active condition, tests were performed on both the affected and the unaffected shoulder to provide a within subject control. The camera angle for these tests were approximately frontal to start. Moderate head motion was common. Video of each trial was rated offline by a FACS certified coder. To assess inter-observer agreement, 1738 frames selected from one affected-side trial and one unaffected-side trial of 20 participants were randomly sampled and independently coded. Intercoder percent agreement as calculated by the Ekman-Friesen formula [6] was 95%, which compares favorably with other research in the FACS literature. For more information on the database, please refer to [17].

Out of the database, we used 203 sequences from 25 different subjects. Of these 203 sequences, 111 were on the affected shoulder and 92 on the unaffected shoulder. Examples of the video sequences are given in Figure 1. As can be seen in these examples, the data consists of patient's moving their heads throughout the sequence. This highlights the challenge of this problem, as registering the patient's face can become difficult due to these added variabilities. The data is also different from a lot of currently available datasets as the video sequences have various durations, with sequences last from 90 to 700 frames. Within these se-



Figure 1. Examples from the UNBC-McMaster database, showing the instances of pain and also of head pose variation during the sequence.

quences, the patient may display various expressions multiple times.

3. AU Recognition System

Our pain recognition system consists of firstly tracking a patient’s face throughout a sequence. We do this using active appearance models (AAMs). Once the face is tracked, we use the information from the AAMs to extract both contour and appearance features. We then use these features as an input into a support vector machine (SVM), which we use for classification. We explain each of these modules in the following subsections.

3.1. Active Appearance Models (AAMs)

Active appearance models (AAMs) have been shown to be a good method of aligning a pre-defined linear shape model that also has linear appearance variation, to a previously unseen source image containing the object of interest. In general, AAMs fit their shape and appearance components through a gradient-descent search, although other optimization methods have been employed with similar results [4].

The shape \mathbf{s} of an AAM [4] is described by a 2D triangulated mesh. In particular, the coordinates of the mesh vertices define the shape \mathbf{s} . These vertex locations correspond to a source appearance image, from which the shape was aligned. Since AAMs allow linear shape variation, the shape \mathbf{s} can be expressed as a base shape \mathbf{s}_0 plus a linear combination of m shape vectors \mathbf{s}_i :

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^m p_i \mathbf{s}_i \quad (1)$$

where the coefficients $\mathbf{p} = (p_1, \dots, p_m)^T$ are the shape parameters. These shape parameters can typically be divided into similarity parameters \mathbf{p}_s and object specific parameters \mathbf{p}_o , such that $\mathbf{p}^T = [\mathbf{p}_s^T, \mathbf{p}_o^T]$. Similarity parameters are associated with geometric similarity transform (i.e. translation, rotation and scale). The object-specific parameters, are the residual parameters representing geometric variations associated with the actual object shape (e.g., mouth opening, eyes shutting, etc.). Procrustes alignment [4] is employed to estimate the base shape \mathbf{s}_0 .

Keyframes within each video sequence were manually labelled, while the remaining frames were automatically aligned using a gradient descent AAM fit described in [13]. Figure 2(a) and (b) shows the AAM in action, with the 68 point mesh being fitted to the patient’s face in every frame.

3.2. Feature Extraction

Once we have tracked the patient’s face using the AAM by estimating the base shape and appearance parameters, we can use this information to derive features from the face. From the initial work conducted in [1, 11], we extracted the following features:

- **PTS:** similarity normalized shape, \mathbf{s}_n , refers to the 68 vertex points in \mathbf{s}_n for both the x - and y - coordinates, resulting in a raw 136 dimensional feature vector. These points are the vertex locations after all the rigid geometric variation (translation, rotation and scale), relative to the base shape, has been removed. The similarity normalized shape \mathbf{s}_n can be obtained by synthesizing a shape instance of \mathbf{s} , using Equation 1, that ignores the similarity parameters \mathbf{p} . An example of the normalized shape features, PTS, is given in Figure 2(c).
- **APP:** canonical normalized appearance \mathbf{a}_0 refers to where all the non-rigid shape variation has been normalized with respect to the base shape \mathbf{s}_0 . This is accomplished by applying a piece-wise affine warp on each triangle patch appearance in the source image so that it aligns with the base face shape. If we can remove all shape variation from an appearance, we’ll get a representation that can be called as shape normalized appearance, \mathbf{a}_0 . The canonical normalized appearance \mathbf{a}_0 is different to the similarity normalized appearance \mathbf{a}_n as it removes the non-rigid shape variation and not the rigid shape variation. In previous work [1], it was shown by removing the rigid shape variation, poor performance was gained. As such, only the canonical normalized appearance features \mathbf{a}_0 were used in this paper. The resulting features yield an approximately 27,000 dimensional raw feature vector. A mask is applied to each image so that the same amount of pixels are used. Obviously, having such a high dimensional vector is prohibitive for any classifier with

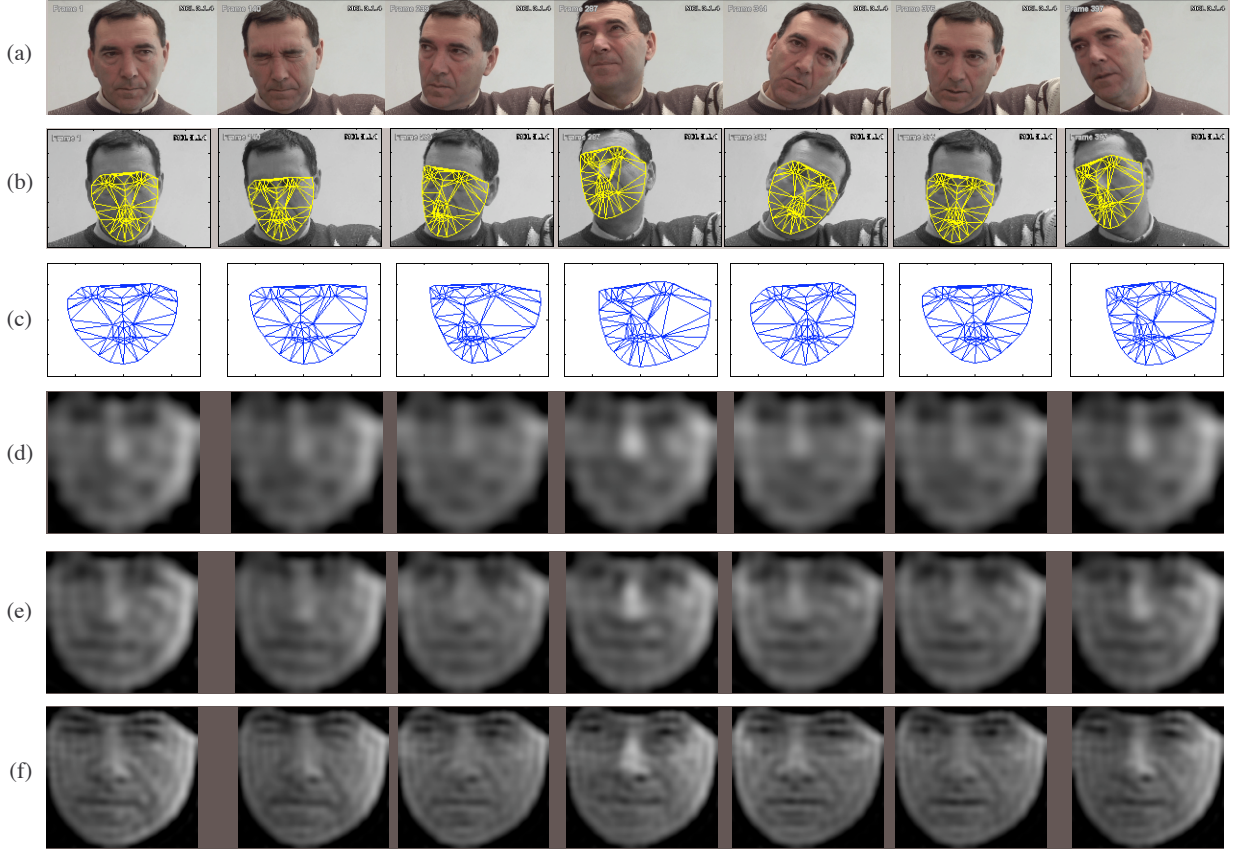


Figure 2. Example of the output of the AAM tracking and the associated shape and appearance features: (a) the original sequence, (b) the AAM tracked sequence, (c) the normalized shape features (PTS), (d) the normalized appearance features using 100 DCT coefficients (APP100), (e) normalized appearance using 200 DCT coefficients (APP200), and (f) normalized appearance using 500 DCT coefficients (APP500).

many training examples, so a compact representation of this frame has to be found. In Ashraf et al. [1], they used K-means clustering in the temporal domain to overcome this constraint. However, in Lucey et al. [10] they found that gaining a compact spatial representation was more advantageous as no temporal information was lost. As such, we used the 2-D discrete cosine transform (DCT). The DCT was used instead of principal component analysis (PCA) as it does not require any prior, as well as being quick to execute (due to the use of the fast fourier transform (FFT)), while achieving the same performance [15]. For this work, we experimented with $M = 100, 200$ and 500 . Examples of the reconstructed images with these coefficients are shown in Figures 2(d, e and f). From these examples, it can be seen the more features used provides a more recognizable reconstruction. It is also worth noting that regardless of the head pose and orientation, the appearance features are projected back onto the normalized base shape, so as to make these features more robust to these visual variabilities.

- **PTS+APP:** combination of the similarity normalized shape and canonical normalized appearance features $s_n + a_0$ refers to the shape features being concatenated to the appearance features. We experimented with this combined feature set with $M = 100, 200$ and 500 coefficients.

3.3. Support Vector Machine Classification

SVMs have been proven useful in a number of pattern recognition tasks including face and facial action recognition. Because they are binary classifiers they are well suited to the task of AU recognition (i.e. AU vs no AU). SVMs attempt to find the hyperplane that maximizes the margin between positive and negative observations for a specified class. A linear SVM classification decision is made for an unlabeled test observation \mathbf{x}^* by,

$$\mathbf{w}^T \mathbf{x}^* \begin{cases} >_{true} & b \\ <_{false} & \end{cases} \quad (2)$$

AU	N	PTS	APP100	APP200	APP500	PTS+APP100	PTS+APP200	PTS+APP500
4	2039	83.51	53.66	60.41	67.94	75.45	76.51	76.33
6	6686	77.08	83.16	82.17	81.91	78.71	81.15	81.41
7	3518	64.68	72.65	72.55	67.91	67.09	61.75	67.67
9	502	74.81	68.81	79.93	70.99	64.98	80.24	77.67
10	624	65.20	79.84	85.56	89.69	76.30	82.78	87.33
12	9040	81.68	76.32	75.70	80.00	79.88	80.04	81.81
20	773	66.03	55.00	56.08	61.25	64.32	64.42	64.47
25	3149	65.58	55.15	58.80	61.94	66.56	66.06	70.44
26	2722	54.73	55.62	58.62	55.46	53.94	52.32	51.33
43	2740	93.17	82.67	82.69	76.38	91.08	88.39	87.19
AVG		75.22	72.01	72.98	73.68	74.68	74.66	76.16

Table 1. Results showing the area underneath the ROC curve for the shape and appearance features. Note the average is a weighted one, depending on the number of positive examples.

where \mathbf{w} is the vector normal to the separating hyperplane and b is the bias. Both \mathbf{w} and b are estimated so that they minimize the structural risk of a train-set, thus avoiding the possibility of overfitting to the training data. Typically, \mathbf{w} is not defined explicitly, but through a linear sum of support vectors. As a result SVMs offer additional appeal as they allow for the employment of non-linear combination functions through the use of kernel functions, such as the radial basis function (RBF), polynomial and sigmoid kernels. A linear kernel was used in our experiments due to its ability to generalize well to unseen data in many pattern recognition tasks [7]. Please refer to [7] for additional information on SVM estimation and kernel selection.

4. Action Unit Recognition

For the facial AU recognition system, we developed a system for each particular feature set (see Section 3.2). The system was designed to determine which AUs are present in each frame of the video sequence. All up, we conducted recognition on ten AUs (4, 6, 7, 9, 10, 12, 20, 25, 26 and 43). To check for subject generalization, a leave-one-subject-out strategy was used and each sub-recognizer was trained using positive examples which consisted of the frames that the FACS coder labelled containing that particular AU (regardless of intensity, i.e. A-E). The negative examples consisted of all the other frames that were not labelled with that particular AU.

In order to predict whether or not a video frame contained an AU, the output score from the SVM was used. As there are many more non-AUs frames than AU frames, the overall agreement between correctly classified frames can skew the results somewhat. As such we used the receiver-operator characteristic (ROC) curve, which is a more reliable performance measure. This curve is obtained by plotting the hit-rate (true positives) against the false alarm rate (false positives) as the decision threshold varies. From the

ROC curves, we used the area under the ROC curve (A'), to assess the performance which has been used in similar studies [9]. The A' metric ranges from 50 (pure chance) to 100 (ideal classification)¹.

4.1. AU Recognition Results

The results for the AU recognition with respect to the shape and appearance features are shown in Table 1. Between the shape (PTS) and appearance (APP) features, it can be seen depending on the AU, the best performing feature set varies. For AUs 4, 12, 20, 25 and 43, the PTS features yielded the higher recognition rates. Conversely, for AUs 6, 7, 9, 10 and 26 the APP features got the better performance, whilst the combined features were not the best performing for any of the AUs.

A explanation of these results can stem from the AAM 2-D mesh that tracks the patient’s face in Figure 2(b). For AU4 (brow lowering), 12 (lip corner pull), 20 (lip stretcher), 25 (lips part) and 43 (eye closing), the areas of the face in which movement pertaining to these AUs occurs lie on the 2-D mesh. So it is intuitive that the most discriminating features for these actions would relate to the shape features. Even though the deformation in the appearance for these AUs may also be informative, it may be confusable by other facial movements. For example in AU43, the PTS features give a clear indication on whether or not the eye was shutting as shown by the very high recognition rate (93.17). This is because the points on the mesh for the eye do not directly correspond to other AUs, so in essence it is somewhat independent of other features. However, for other AUs such as 6 (cheek raiser) and 7 (lid tightener), these facial movements are associated with the skin deforming around the eye causing wrinkles which can be seen by a change in the appearance around these areas. This suggests why the

¹In literature, the A' metric varies from 0.5 to 1, but for this work we have multiplied the metric by 100 for improved readability of results

appearance features are more discriminating for these features.

Another interesting result emanating from these experiments is the effect the number of DCT coefficients had on the appearance features. It would be intuitive that increasing the number of features would in turn improve the performance due the added information included. Whilst this was the case for AUs 4, 10, 12, 20 and 25, the trend in the other AUs is somewhat random. This randomness maybe due noise introduced by the synthesis of the normalized appearance. Overall however, the weighted averages indicate that when more features are used, better performance is achieved, albeit by a small amount (i.e. APP100 = 72.01, APP200 = 72.98 and APP500 = 73.68).

As eluded to previously, the combined representations did not outperform either the PTS or APP features for any of the AUs. Even though the combined feature set using 500 appearance features (PTS+APP500) gained the best performance out of all feature sets with an weighted average of $A' = 76.16$, this can be somewhat misleading as the weighted average of the best performing features for each AU outperforms this mark by achieving $A' = 78.03$.

For clearer analysis we provided the ROC curves for each of the AUs in Figure 3 (only the APP500 and the PTS+APP500 are shown also AU20, 25 and 26 were omitted). From these curves it can be seen that a very good level of performance was achieved by AUs 4,12 and 43 for the PTS features whilst the appearance features gained better performance for AU6 and 10. It is hard to tell which feature set was better for AU 7 and 9 as they did not perform that well ($A' \leq 70$), however in the case of AU9 there were not that many examples which may have effected the performance. But again, the combination of the APP and the PTS features seems to have just averaged out the performance between the two sets of data and not provide any complementary information. These results adds weight to our argument for using different features for recognizing different AUs.

For clearer analysis we provided the ROC curves for each of the AUs in Figure 3 (only the APP500 and the PTS+APP500 are shown and AU20, 25 and 26 were omitted). From these curves it can be seen that a very good level of performance was achieved by AUs 4, 12 and 43 for the PTS features whilst the appearance features gained better performance for AU6 and 10. It is hard to tell which feature set was better for AU 7 and 9 as they did not perform that well ($A' < 70$), however in the case of AU9 there were not that many examples which may have effected the performance. For AUs 4, 6, 10 and 43 the combination of both shape and appearance decreases accuracy as the APP and the PTS features seems to have just averaged out the performance between the two sets of data. These results adds weight to our argument for using different features for rec-

ognizing different AUs.

5. Conclusions and Future Work

We conducted a series of experiments to determine whether there was benefit in using different feature representations for the detection of specific facial action units (AUs). Most current systems to automatically recognize AUs use appearance. Based on the experience of manual FACS coders and knowledge of how facial features deform with movement, we hypothesized that shape and appearance would differentially contribute to the detection of specific action units. Specifically, action units such as brow lowering (AU4) would be better detected by shape features; whereas action units such as cheek raising (AU6) would be better detected by appearance. We found strong support for this hypothesis. Experiments were conducted using the UNBC-McMaster Shoulder Pain Archive, which contains spontaneous facial actions of subjects undergoing assessment for shoulder-related pain. This database represents the type of real-world application for which automatic AU detection is needed.

Shape yielded the best results for AU4 and AU43. Both features are well represented by deformations of the AAM mesh. The addition of appearance features for these AU only served to decrease accuracy. We suspect that one reason for this finding is the large interpersonal variability in appearance associated with brow lowering and eye closure, especially when these actions are strong. For several AUs, best results were obtained by using appearance. Examples were AU6 and AU10. For these AUs, the addition of appearance features may have served to reduce confusability when compared to the shape domain. Finally, in at least one case, AU12, shape and appearance gave comparable results. In summary, automatic AU detection in ecologically valid settings appears feasible. Accuracy is highest when shape and appearance features are selected with respect to the specific AU targets.

6. Acknowledgements

This project was supported in part by CIHR Operating Grant MOP77799 and National Institute of Mental Health grant R01 MH51435. Zara Ambadar, Nicole Grochowina, Amy Johnson, David Nordstokke, Nicole Ridgeway, and Nathan Unger provided technical assistance.

References

- [1] A. Ashraf, S. Lucey, J. Cohn, T. Chen, Z. Ambadar, K. Prkachin, P. . Solomon, and B.-J. Theobald. The painful face: pain expression recognition using active appearance models. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 9–14. ACM, Nagoya, Aichi, Japan, 2007. 2, 3, 4
- [2] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. In *Proceedings of the Interna-*

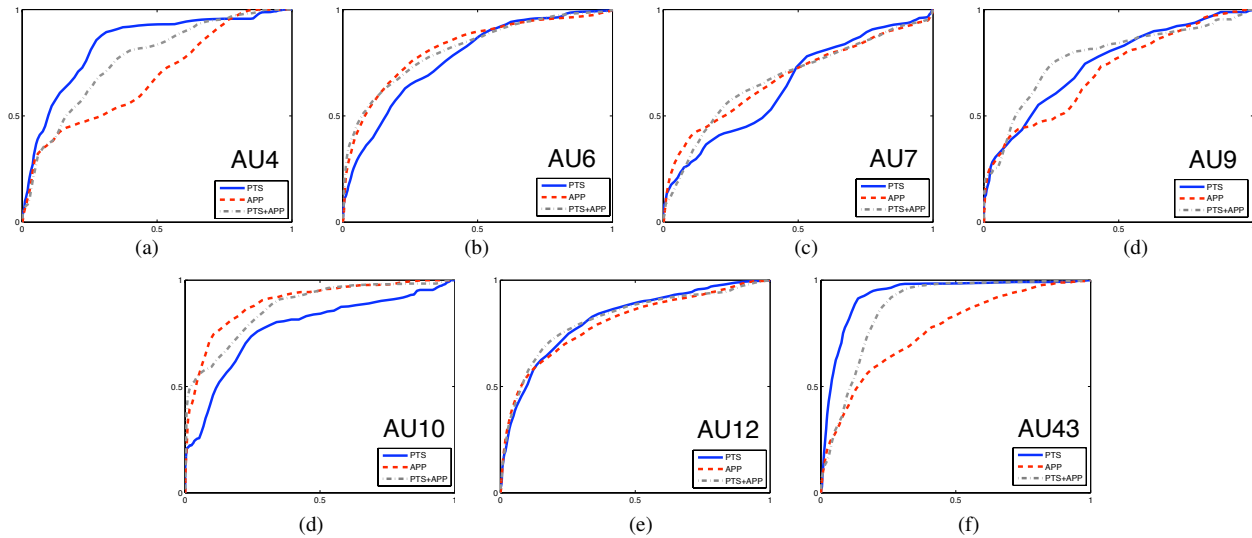


Figure 3. ROC curve for selected AUs. The solid line is the PTS features, the dashed line is the APP features and the dotted line is the PTS+APP features. For the appearance features only the 500 DCT features is depicted

tional Conference on Automatic Face and Gesture Recognition, pages 223–228, 2006. 2

- [3] J. Cohn and K. Schmidt. The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 2:1–12, 2004. 2
- [4] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001. 1, 3
- [5] K. Craig, K. Prkachin, and R. Grunau. The facial expression of pain. In *Handbook of pain assessment*. 1
- [6] P. Ekman, W. Friesen, and J. Hager. *Facial Action Coding System: Research Nexus*. Network Research Information, Salt Lake City, UT, USA, 2002. 1, 2
- [7] C. Hsu, C. C. Chang, and C. J. Lin. A practical guide to support vector classification. Technical report, 2005. 5
- [8] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. *Journal of Image and Vision Computing*, 24(6):615–625. 2
- [9] G. C. Littlewort, M. S. Bartlett, and K. Lee. Faces of pain: automated measurement of spontaneous all facial expressions of genuine and posed pain. In *Proceedings of the International Conference on Multimodal Interfaces*, pages 15–21, New York, NY, USA, 2007. ACM. 5
- [10] P. Lucey, J. Howlett, J. Cohn, S. Lucey, S. Sridharan, and Z. Ambadar. Improving pain recognition through better utilisation of temporal information. In *Proceedings of the International Conference on Auditory-Visual Speech Processing*, Tangalooma, Queensland, Australia, 2008. 4
- [11] S. Lucey, A. Ashraf, and J. Cohn. Investigating spontaneous facial action recognition through aam representations of the face. In K. Kurihara, editor, *Face Recognition Book*. Pro Literatur Verlag, 2007. 2, 3
- [12] S. Lucey, I. Matthews, C. Hu, Z. A. F. de la Torre, and J. Cohn. Aam derived face representations for robust facial action recognition. In I. Matthews, editor, *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 155–160, 2006. 2
- [13] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004. 1, 3
- [14] M. Pantic and L. Rothkrantz. Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 34(3):1449–1461, 2004. 2
- [15] G. Potamianos, C. Neti, G. Iyengar, A. Senior, and A. Verma. A cascade visual front end for speaker independent automatic speechreading. *International Journal of Speech Technology*, 4(3-4):193–208, 2001. 4
- [16] K. Prkachin. The consistency of facial expressions of pain: a comparison across modalities. *Pain*, 51:297–306, 1992. 1
- [17] K. Prkachin and P. Solomon. The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. *Pain*, 139:267–274, 2008. 1, 2
- [18] Y. Tian, J. Cohn, and T. Kanade. Facial expression analysis. In S. Li and A. Jain, editors, *The handbook of emotion elicitation and assessment*, pages 247–276. Springer, New York, NY, USA. 2
- [19] Y. Tong, W. Liao, and Q. Ji. Facial action unit recognition by exploiting their dynamic and semantic relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10):1683–1699, 2007. 2
- [20] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518, 2001. 2
- [21] A. Williams, H. Davies, and Y. Chadury. Simple pain rating scales hide complex idiosyncratic meanings. *Pain*, 85:457–463, 2000. 1