21

Investigating Spontaneous Facial Action Recognition through AAM Representations of the Face

Simon Lucey, Ahmed Bilal Ashraf, Jeffrey F. Cohn
Carnegie Mellon University
USA

1 Introduction

The Facial Action Coding System (FACS) [Ekman et al., 2002] is the leading method for measuring facial movement in behavioral science. FACS has been successfully applied, but not limited to, identifying the differences between simulated and genuine pain, differences betweenwhen people are telling the truth versus lying, and differences between suicidal and non-suicidal patients [Ekman and Rosenberg, 2005]. Successfully recognizing facial actions is recognized as one of the "major" hurdles to overcome, for successful automated expression recognition.

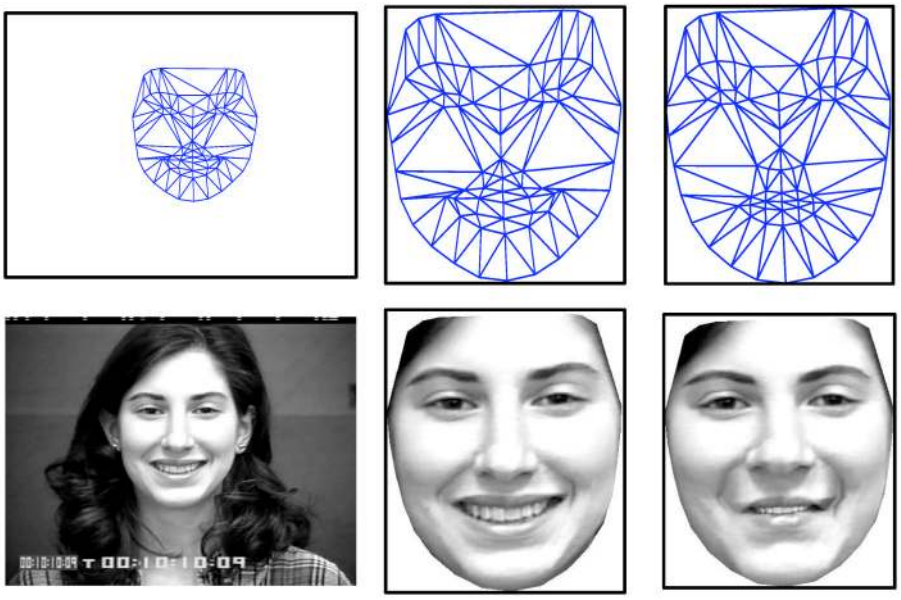
How one should represent the face for effective action unit recognition is the main topic of interest in this chapter. This interest is motivated by the plethora of work in existence in other areas of face analysis, such as face recognition [Zhao et al., 2003], that demonstrate the benefit of representation when performing recognition tasks. It is well understood in the field of statistical pattern recognition [Duda et al., 2001] given a fixed classifier and training set that how one represents a pattern can greatly effect recognition performance. The face can be represented in a myriad of ways. Much work in facial action recognition has centered solely on the appearance (i.e., pixel values) of the face given quite a basic alignment (e.g., eyes and nose). In our work we investigate the employment of the Active Appearance Model (AAM) framework [Cootes et al., 2001, Matthews and Baker, 2004] in order to derive effective representations for facial action recognition. Some of the representations we will be employing can be seen in Figure 1.

Experiments in this chapter are run across two action unit databases. The Cohn- Kanade FACS-Coded Facial Expression Database [Kanade et al., 2000] is employed to investigate the effect of face representation on *posed* facial action unit recognition. Posed facial actions are those that have been elicited by asking subjects to deliberately make specific facial actions or expressions. Facial actions are typically recorded under controlled circumstances that include full-face frontal view, good lighting, constrained head movement and selectivity in terms of the type and magnitude of facial actions. Almost all work in automatic facial expression analysis has used posed image data and the Cohn-Kanade database may be the database most widely used [Tian et al., 2005]. The RU-FACS Spontaneous Expression Database is employed to investigate how these same representations affect *spontaneous* facial action unit recognition. Spontaneous facial actions are representative of "real-world" facial

actions. They typically occur in less controlled environments, with non-frontal pose, smaller face image size, small to moderate head movement, and less intense and often more complex facial actions. Spontaneous actions units are elicited indirectly from subjects through environmental variables (e.g., showing a subject something associated with happiness which then indirectly results in a smile). Although harder to collect and annotate, spontaneous facial actions are preferable to posed as they are representative of real world facial actions. Most automated facial action recognition systems have only been evaluated on posed facial action data [Donato et al., 1999, Tian et al., 2001] with only a small number of studies being conducted on spontaneous data [Braathen et al., 2001, Bartlett et al., 2005].
This study extends much of the earlier work we conducted in [Lucey et al., 2006]. We greatly expand upon our earlier work in a number of ways. First, we expand the number of AUs analyzed from 4, centered around the brow region, to 15, stemming from all regions of the face. Second, we investigate how representation affects both posed and spontaneous actions units by running our experiments across both kinds of datasets. Third, we report results in terms of verification performance (i.e., accept or reject that a claimed AU observation is that AU) rather than identification performance (i.e., determine out of a watchlist of AU combinations which class does this observation belong to?). The verification paradigm is preferable over identification as it provides a natural mechanism for dealing with simultaneously occurring AUs and is consistent with existing literature [Bartlett et al., 2005].

1.1 Background

One of the first studies into representations of the face, for automatic facial action recognition, was conducted by Donato et al. [1999]. Motivated by the plethora of work previously performed in the face recognition community, this study was restricted to 2-D appearance based representations of the face (e.g. raw pixels, optical flow, Gabor filters, etc.) as well as data-driven approaches for obtaining compact features (e.g. PCA, LDA, ICA, etc.). These appearance based approaches were broadly categorized into holistic, also referred to as monolithic, and parts-based representations. In the ensuing literature, appearance based approaches have continued to be popular as demonstrated by the recent feature evaluation paper by Bartlett et al. [2005]. A major criticism of purely appearance based approaches however, is their lack of shape registration. When "realworld" variation occurs, their lack of shape registration (i.e. knowledge of the position of the eyes, brow, mouth, etc.) makes normalizing for translation and rotation difficult to achieve.
Model-based approaches offer an alternative to appearance based approaches for representing the face. Typical approaches have beenActive ShapeModels (ASMs) [Cootes et al., 2001] and Active Appearance Models (AAMs) [Cootes et al., 2001, Matthews and Baker, 2004] in which both appearance and shape can be extracted and decoupled from one another. Model-based approaches to obtaining representations, like those possible with AAMs, have an inherent benefit over purely appearance based representations in the sense they can account and attempt to normalize for many types of "real-world" variation.



(a) (b) (c)

Figure 1. This figure depicts the different levels of shape removal from the appearance. Column (a) depicts the initial scenario in which all shape and appearance is preserved. In (b) geometric similarity is removed from both the shape and appearance; and in (c) shape (including similarity) has been removed leaving the *average* face shape and what we refer to as the face image's *canonical appearance*. Features derived from the representations in columns (b) and (c) are used in this paper for the task of facial action unit recognition

1.2 Scope

Contraction of the facialmuscles produces changes in the appearance and shape of facial landmarks (e.g., brows) and in the direction and magnitude of motion on the surface of the skin and in the appearance of transient facial features (e.g., wrinkles). It is due to the differing nature of these changes in face shape and appearance that we hypothesize that AAM derived representations could be beneficial to the task of automatic facial action recognition.

The scope of this paper was restricted to the specific task of peak versus peak AU verification. Typically, when an AU is annotated there may be a time stamp noted for its onset (i.e., start), offset (stop) and/or peak (maximum intensity). For the Cohn-Kanade database, time stamps were provided for onset and peak AU intensity of each image sequence. For RU-FACS, time stamps were available for onset, peak, and offset. For both databases, AUs typically were graded for intensity, with A being the lowest grade intensity (i.e. "trace" or barely visible; not coded in the original edition of FACS) and E being the highest [Ekman et al., 2002]. Only AUs of at least intensity B were employed in our experiments. Onset time stamps were assumed to be representative of a local AU 0 (i.e. neutral expression). AU 0 is employed later in our experiments in a normalization technique.

The reliability of annotated action units was considered in selecting image sequences for analysis. If manual FACS coding is contaminated by error, the potential verification rate is proportionately reduced. For the Cohn-Kanade database, reliability of annotated AU was evaluated by independently scoring a random subset of the image sequences. Reliability for AU occurrence was moderate to high [Kanade et al., 2000]. We therefore used all available image sequences for analysis. For RU-FACS, reliability is not reported. In its absence, a certified FACS coder fromthe University of Pittsburgh verified all action units. Sequences for which manual FACS coding was not confirmed were excluded.

2. AAMs

In this section we describe active appearance models (AAMs). AAMs have been demonstrated to be an excellentmethod for aligning a pre-defined linear shapemodel, that also has linear appearance variation, to a previously unseen source image containing that object. AAMs typically fit their shape and appearance components through a gradientdescent fit, although other optimization approaches have been employed with similar results [Cootes et al., 2001]. To ensure a quality fit, for the datasets employed in this study, a subject-dependent AAM was created for each subject in each database. Keyframes taken from each subject were manually labeled in order to create the subjectdependent AAM. The residual frames for the subject were then aligned in an automated fashion using a gradient-descent AAM fit. Please see [Matthews and Baker, 2004] for more details on this approach.

2.1 AAM Derived Representations

The *shape* $\mathbf{s}$ of an AAM [Cootes et al., 2001] is a 2D triangulated mesh. In particular, $\mathbf{s}$ is a column vector containing the vertex locations of the mesh (see row 1, column (a), of Figure 1 for examples of this mesh). These vertex locations correspond to a source appearance image, fromwhich the shapewas aligned (see row 2, column (a), of Figure 1).

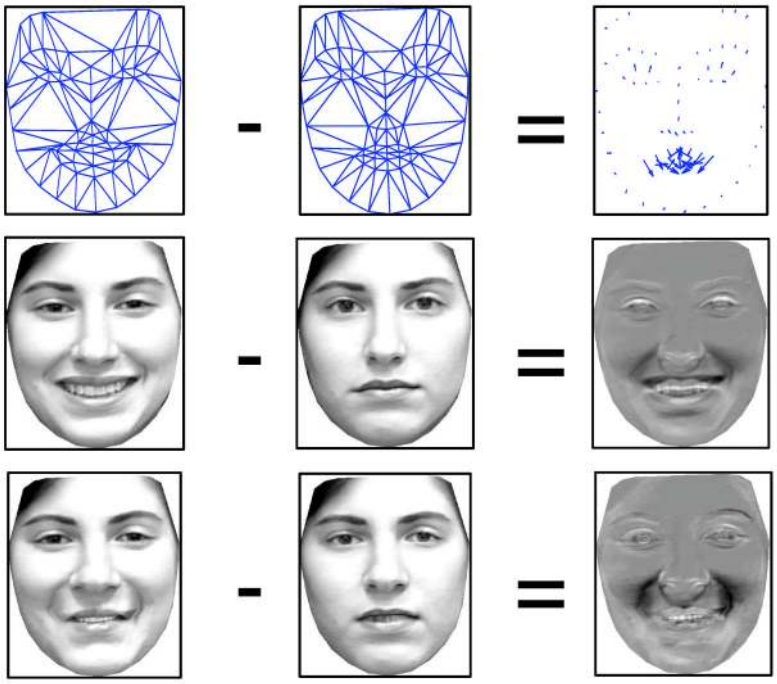
AAMs allow linear shape variation. This means that the shape $\mathbf{s}$ can be expressed as a base shape $\mathbf{s}_0$ plus a linear combination of m shape vectors $\mathbf{s}_i$:

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{m} p_i \, \mathbf{s}_i \tag{1}$$

where the coefficients $\mathbf{p} = (p_1, \ldots, p_m)^{\mathrm{T}}$ are the shape parameters. These shape parameters can typically be divided into similarity parameters $\mathbf{p}_s$ and object-specific parameters $\mathbf{p}_0$, such that $\mathbf{p}^T = [\mathbf{p}_s T, \ \mathbf{p}_0{}^T]$. Similarity parameters are associated with the geometric similarity transform (i.e., translation, rotation and scale). The object-specific parameters, are the residual parameters associated with geometric variations associated with the actual object shape (e.g., the mouth opening, eyes shutting, etc.). Procrustes alignment [Cootes et al., 2001] is employed to estimate the base shape $\mathbf{s}_0$.

A 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 of $\mathbf{p}$. An example of this similarity normalized mesh can seen in row 1, column (b), of Figure 1. A similarity normalized appearance can then be synthesized by employing a piece-wise affine warp on each triangle patch appearance in the source image (see row2, column (b), of Figure 1) so the appearance contained within $\mathbf{s}$ now aligns with the similarity normalized shape $\mathbf{s}_n$. We shall refer to this

as the face's *similarity normalized appearance* $\mathbf{a}_n$. A shape normalized appearance can then be synthesized by applying the same technique, but instead ensuring the appearance contained within $\mathbf{s}$ now aligns with the base shape $\mathbf{s}_0$. We shall refer to this as the face's *canonical appearance* (see row 2, column (c), of Figure 1 for an example of this canonical appearance image) $\mathbf{a}0$.



(a) (b) (c)

Figure 2. This figure depicts a visualization of *delta features* for **S-PTS** (row 1), **S-APP** (row 2) and **C-APP** (row 3). The peak and neutral frames for these different features can be seen in columns (a) and (b) respectively. The delta features can be seen in column (c)

2.2 Features

Based on the AAM derived representations in Section 2.1 we define three representations:

S-PTS: *similarity normalized shape* $\mathbf{s}_n$ representation (see Equation 1) of the face and its facial features. There are 74 vertex points in $\mathbf{s}_n$ for both x- and y- coordinates, resulting in a raw 148 dimensional feature vector.

S-APP: *similarity normalized appearance* $\mathbf{a}_n$ representation. Due to the number of pixels in $\mathbf{a}_n$ varying from image to image, we apply a mask based on $\mathbf{s}_0$ so that the same number of pixels (approximately 126, 000) are in $\mathbf{a}_n$ for each image.

C-APP: *canonical appearance* $\mathbf{a}0$ representation where all shape variation has been removed from the source appearance except the base shape $\mathbf{s}_0$. This results in an approximately 126, 000 dimensional raw feature vector based on the pixel values within $\mathbf{s}_0$.

The naming convention **S-PTS**, **S-APP**and **C-APP** will be employed throughout the rest of this chapter.

In previous work [Cohn et al., 1999, Lucey et al., 2006] it has been demonstrated that some formof subject normalization is beneficial in terms of recognition performance. The employment of *delta features* are a particularly useful method for subject normalization (see Figure 2). A delta feature $\mathbf{x}_?$ is defined as,

$$\mathbf{x}_? = \mathbf{x}_{peak} - \mathbf{x}_{neutral} \tag{2}$$

where $\mathbf{x}_{peak}$ is the feature vector taken at the peak time stamp for the current AU being verified. The $\mathbf{x}_{neutral}$ feature vector is taken from a neutral time stamp for that same subject. The feature $\mathbf{x}$ is just notation for any generic feature, whether it stem from **S- PTS S-APP** or **C-APP**. The employment of delta features is advantageous as it can lessen the effect of subject-dependent bias during verification. A visualization of delta features can be seen in Figure 2 for **S-PTS** (row 1), **S-APP** (row 2) and **C-APP** (row 3).

3. Classifiers

Becausewe are dealingwith peak-to-peakAUverification,we explored two commonly [Donato et al., 1999, Bartlett et al., 2005] used classifiers for facial action recognition.

3.1 Support Vector Machine (SVM)

Support vector machines (SVMs) have been demonstrated to be extremely useful in a number of pattern recognition tasks including face and facial action recognition. This type of classifier attempts to find the hyper-plane 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,

$$\mathbf{w}^T\mathbf{x}^* \underset{\text{false}}{\overset{\text{true}}{\gtrless}} b \tag{3}$$

where $\mathbf{w}$ is the vector normal to the separating hyperplane and b is the bias. Both $\mathbf{w}$ and b were estimated so that they minimize the structural risk of a train-set. 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*, *sigmoid* kernels. A linear kernel was used in our experiments throughout this chapter, however, due to its good performance, and ability to perform well in many pattern recognition tasks Hsu et al. [2005]. Please refer to [Hsu et al., 2005] for additional information on SVM estimation and kernel selection.

Since SVMs are intrinsically binary classifiers, special steps must be taken to extend them to the multi-class scenario required for facial action recognition. In our work, we adhered to the "one-against-one" approach [Hsu et al., 2005] in which $K(K - 1)/2$ classifiers are constructed, where K are the number of AU classes, and each one trains data from two different classes. In classification we use a voting strategy, where each binary classification is considered to be a single vote. A classification decision is achieved by choosing the class with the maximum number of votes.

3.2 Nearest Neighbor (NN)

Nearest neighbor (NN) classifiers are typically employed in scenarios where there are many classes, and there is a minimal amount of training observations for each class (e.g. face recognition); making them well suited for the task of facial action recognition.

A NN classifier seeks to find of N labeled train observations $\{\mathbf{x}_i\}_{i=1}^{N}$ the single closest observation to the unlabeled test observation $\mathbf{x}^*$; classifying $\mathbf{x}^*$ as having the nearest neighbor's label.

AU	N	P	FAR	FRR
1	141	93.35	20.57	0.88
2	94	97.09	7.45	1.81
4	154	88.98	28.57	2.75
5	77	93.35	16.88	4.70
6	111	88.98	24.32	7.03
7	108	88.98	34.26	4.29
9	50	98.75	10.00	0.23
12	113	96.88	7.08	1.90
15	73	96.88	13.70	1.23
17	153	95.63	3.92	4.57
20	69	94.59	33.33	0.73
23	43	95.01	44.19	1.14
24	43	95.84	41.86	0.46
25	293	98.13	3.07	0.00
27	76	97.30	3.95	2.47
Average		***94.65***	***19.54***	***2.28***

Table 1. Verification results for similarity removed geometric shape **S-PTS** features. Results are depicted for each action unit (AU), with the number of positive examples (N), total percentage agreement between positive and negative labels (P), false accept rate (FAR), false reject rate (FRR)

When N is small the choice of distancemetric $D(\mathbf{a}, \mathbf{b})$ between observation points becomes especially important [Fukunaga, 1990]. One of themost common distancemetrics employed in face recognition and facial action recognition is the Mahalanobis distance,

$$D(\mathbf{a}, \mathbf{b}) = (\mathbf{a} - \mathbf{b})^T \mathbf{W} (\mathbf{a} - \mathbf{b}) \tag{4}$$

where $\mathbf{a}$ and $\mathbf{b}$ are observation vectors being compared and $\mathbf{W}$ is a weightingmatrix. It is often advantageous to attempt to learn $\mathbf{W}$ from the train-set. Two common approaches to learn $\mathbf{W}$ are,

Principal Component Analysis (PCA) attempts to find the K eigenvectors $\mathbf{V} = \{\mathbf{v}_k\}_{k=1}^{K}$, corresponding to the K largest eigenvalues, of the train-set's covariance matrix. These K eigenvectors can be thought of as the K largest modes of linear variation in the train-set. The weighting matrix can then be defined as $\mathbf{W} = \mathbf{V}\mathbf{V}^T$. Typically, $K << N$ thereby constraining the matching of $\mathbf{a}$ and $\mathbf{b}$ to a subspace where training observations have previously spanned.

Linear Discriminant Analysis (LDA) attempts to find the K eigenvectors $\mathbf{V} = \{\mathbf{v}_k\}_{k=1}^{K}$ of $\mathbf{S}_b\mathbf{S}_w^{-1}$ where $\mathbf{S}_b$ and $\mathbf{S}_w$ are the within- and between- class scatter matrices of the train-set. These K eigenvectors can be thought of as the K largest modes of discrimination in the train-set. Since $\mathbf{S}_b\mathbf{S}_w^{-1}$ is not symmetrical, we must employ simultaneous diagonalization Fukunaga [1990] to find the solution. PCA is typically applied before LDA, especially if the dimensionality of the raw face representations is large, so as to minimize sample-size noise.

In our initial experiments we found no advantage in employing NN classifiers, based on either PCA or LDA subspaces, when compared to SVM classifiers. This result was consistent with our own previous work [Lucey et al., 2006] and other previous work in literature [Bartlett et al., 2005]. In the interests of succinctness we shall only be reporting verification results for SVM classifiers.

4 Experiments

4.1 Evaluation

Verification is performed by accepting a claimed action unit when its match-score is greater than or equal to *Th* and rejecting the claimed action unit when the match-score is less than *Th*, where *Th* is a given threshold. Verification performance is evaluated using twomeasures; being false rejection rate (FRR), where a true action unit is rejected against their own claim, and false acceptance rate (FAR), where an action unit is accepted as the falsely claimed action unit. The FAR and FRR measures increase or decrease in contrast to each other based on the threshold *Th*. In our experiments an optimized threshold *Th** was learnt in conjunction with the SVM classifier that minimizes the *total* number of falsely classified training observations.

4.2 Posed Action Units

In our first set of experimentswe investigated howdifferent representations affected verification performance of a "posed" set of action units. The set of AUs employed for our verification performance were based off previous verification experiments conducted by Bartlett et al. [2004]. Verification results can be seen in Tables 1-3. We employed the Cohn-Kanade database for our experiments on posed action units. Due to the small size of the databases being employed for our evaluation, we employed a subject leave-one-out strategy [Duda et al., 2001] to maximize the amount of available training data.

One can see that all three representations achieve reasonable verification performance in terms of FAR, FRR as well as the overall agreement in class labels for both types of error (P). Interestingly, however, the **S-PTS+C-APP** features in Table 3 obtain the best verification performance overall in comparison with the similarity normalized shape (**S-PTS**, Table 1) and appearance (**S-APP**, Table 2) features. The **S-PTS+C-APP** features are created by concatenating together the similarity normalized shape and the shape normalized (canonical) appearance. This result is intriguing as the **S-APP** features contain exactly the same information as the **S-PTS+C-APP** features. The major difference between the two results lies solely in the representation employed. This results demonstrates some of the inherent advantages in employing AAM based representations for facial action unit recognition.

AU	N	P	FAR	FRR
1	141	93.56	19.86	0.88
2	94	96.47	10.64	1.81
4	154	92.31	16.88	3.36
5	77	92.10	28.57	3.96
6	111	88.98	23.42	7.30
7	108	89.61	32.41	4.02
9	50	98.75	12.00	0.00
12	113	97.30	7.96	1.09
15	73	95.63	24.66	0.74
17	153	96.05	5.23	3.35
20	69	96.05	27.54	0.00
23	43	94.59	51.16	0.91
24	43	95.22	48.84	0.46
25	293	95.22	4.78	4.79
27	76	97.71	10.53	0.74
Averag		***94.64***	***21.63***	***2.23***

Table 2. Verification results for similarity removed appearance **S-APP** features. Results are depicted for each action unit (AU), with the number of positive examples (N), total percentage agreement between positive and negative labels (P), false accept rate (FAR), false reject rate (FRR)

AU	N	P	FAR	FRR
1	141	95.43	14.18	0.59
2	94	96.26	10.64	2.07
4	54	91.68	21.43	2.14
5	77	94.18	19.48	3.22
6	111	91.06	20.72	5.41
7	108	90.44	28.70	4.02
9	50	98.75	10.00	0.23
12	113	97.09	7.08	1.63
15	73	97.51	10.96	0.98
17	153	96.26	3.92	3.66
20	69	95.84	26.09	0.49
23	43	95.84	37.21	0.91
24	43	96.67	30.23	0.68
25	293	98.34	2.73	0.00
27	76	97.09	6.58	2.22
Average		***95.50***	***16.66***	***1.88***

Table 3. Verification results for joint **S-PTS+C-APP** features. For these experiments the **S-PTS** and **C-APP** features were concatenated into a single feature vector. Results are depicted for each action unit (AU), with the number of positive examples (N), total percentage agreement between positive and negative labels (P), false accept rate (FAR), false reject rate (FRR)

		Observed			
		1	1+2	4	5
Actual	1	86.42	11.54	3.85	0.00
	1+2	3.45	96.55	0.00	0.00
	4	12.50	0.00	84.38	3.12
	5	43.75	6.25	18.75	31.25

Table 4. Confusion matrix for the similarity normalized shape feature **S-PTS**, demonstrating good performance on AUs 1, 1+2 and 4, but poor performance on AU 5

When we compare these results to other verification experiments conducted in literature for facial action verification, most notably Bartlett et al. [2004] where experiments were carried out on the same database with the same set of AUs, our approach demonstrates improvement. Our algorithm compares favorably to their approach which reports a FAR = 2.55% and a FRR = 33.06% compared to our leading verification performance of FAR = 1.88% and a FRR = 16.66%. In both approaches a SVM was employed for classification with a subject leave-one-out strategy and a threshold Th^* was chosen that minimizes the total number of falsely classified training observations. Bartlett et al.'s approach differed significantly to our own as they employed Gabor filtered appearance features that were then refined through a Adaboost feature selection process. We must note, however, there were slight discrepancies in the number of observations for each AU class which may also account for our differing performance.

4.3 Spontaneous Action Units

In our next set of experiments we investigated how AAM representations performed on "spontaneous" action units. At the time of publishing only a small number of AUs within the RU-FACS database were confirmed so we limited our spontaneous experiments to only the task of AU identification. In our experiments, we focus on two types of muscle action. Contraction of the frontalis muscle raises the brows in an arch-like shape (AU 1 in FACS) and produces horizontal furrows in the forehead (AU 1+2 in FACS). Contraction of the corrugator supercilii and depressor supercilii muscles draws the inner (i.e., medial) portion of the brows together and downward and causes vertical wrinkles to form or deepen between the brows (AU 4 in FACS). The levator palpebrae superioris (AU 5 in FACS) is associated with the raising of the upper eyelid. Because these action units and action unit combinations in the eye and brow region occur frequently during conversation and in expression of emotion, we concentrated on them in our experiments.

In Table 4 we see the confusion matrix for the representation **S-PTS**. Interestingly the performance of the recognizer suffers mainly from the poor job it does on AU 5. Inspecting Table 5, however, for the **S-APP** appearance feature one can see this recognizer does a good job on AUs 1, 1+2 and 4, but does a better job on AU5 than **S-PTS** does. This may indicate that shape and appearance representations of the face may hold some complimentary information with regard to recognizing facial actions. The **S-PTS+C-APP** features were omitted from this evaluation as we wanted to evaluate the the advantages of shape versus appearance based representations.

		Observed			
		1	1+2	4	5
Actual	1	76.92	19.23	3.85	0
	1+2	13.79	86.21	0	0
	4	15.62	18.75	62.5	3.12
	5	18.75	12.5	12.5	56.25

Table 5. Confusion matrix for the appearance feature 2DA, demonstrating reasonable performance on AUs 1, 1+2 and 4, but much better performance, with respect to **S-APP**, on AU 5

5 Discussion

In this paper we have explored a number of representations of the face, derived from AAMs, for the purpose of facial action recognition. We have demonstrated that a number of representations derived from the AAM are highly useful for the task of facial action recognition. A number of outcomes came from our experiments,

- Employing a concatenated similarity normalized shape and shape normalized (canonical) appearance (**S-PTS+C-APP**) is superior to either similarity normalized shape (**S-PTS**) or similarity normalized appearance (**S-APP**). This result also validates the employment of AAM type representations in facial action unit recognition.
- Comparable verification performance to [Bartlett et al., 2004] can be achieved using appearance and shape features stemming from a AAM representation.
- Shape features have a large role to play in facial action unit recognition. Based on our initial experiments the ability to successfully register the shape of the face can be highly beneficial in terms of AU recognition performance.

A major problem with the "spontaneous" action unit recognition component of this chapter stems from the marked amount of head movement in subjects. Additional work still needs to be done, with model based representations of the face, in obtaining adequate 3D depth information from the face. We believe further improvement in this aspect of model based representations of the face, could play large dividends towards the lofty goal of automatic ubiquitous facial action recognition.

6. References

M. Bartlett, G. Littlewort, C. Lainscsek, I. Fasel, and J.Movellan. Machine learningmethods for fully automatic recognition of facial expressions and facial actions. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 592–597, October 2004.

M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, and J. I. Fasel, I.; Movellan. Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 568–573, June 2005.

B. Braathen, M. S. Bartlett, G. Littlewort, and J. R. Movellan. First steps towards automatic recognition of spontaneous facial action units. In *ACM Conference on Perceptual User Interfaces*, 2001.

J. F. Cohn, A. Zlochower, J. Lien, and T. Kanade. Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology*, 36:35–43, 1999.

T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *PAMI*, 23(6):681–685, 2001.

G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying Facial Actions. *IEEE Trans. PAMI*, 21(10):979–984, October 1999.

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons, Inc., New York, NY, USA, 2nd edition, 2001.

P. Ekman and E. Rosenberg. *What the face reveals*. Oxford, New York, 2nd edition, 2005.

P. Ekman, W. V. Friesen, and J. C. Hager. Facial action coding system. Technical report, *Research Nexus*, Network Research Information, Salt Lake City, UT, 2002.

K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 2nd edition, 1990.

C. W. Hsu, C. C. Chang, and C. J. Lin. A practical guide to support vector classification. *Technical report*, Department of Computer Science, National Taiwan University, 2005.

T. Kanade, J. F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *IEEE International Conference on Automatic Face and Gesture Recognition (AFGR)*, pages 46–53, 2000.

S. Lucey, I.Matthews, C. Hu, Z. Ambadar, F. De la Torre, and J. Cohn. AAMderived face representations for robust facial action recognition. In *IEEE International Conference on Automatic Face and Gesture Recognition (AFGR)*, pages 155–160, 2006.

I. Matthews and S. Baker. Active Appearance Models revisited. *IJCV*, 60(2):135–164, 2004.

Y. Tian, T. Kanade, and J. Cohn. Recognizing action units of facial expression analysis. *IEEE Trans. on PAMI*, 23:229–234, 2001.

Y. Tian, J. F. Cohn, and T. Kanade. Facial expression analysis. In S. Z. Li and A. K. Jain, editors, *Handbook of face recognition*, pages 247–276. Springer, New York, 2005.

W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys (CSUR)*, 35(4):399–458, December 2003.