The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression

Patrick Lucey[1,2], Jeffrey F. Cohn[1,2], Takeo Kanade[1], Jason Saragih[1], Zara Ambadar[2]
Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 15213[1]
Department of Psychology, University of Pittsburgh, Pittsburgh, PA, 15260[2]
patlucey@andrew.cmu.edu, jeffcohn@cs.cmu.edu, tk@cs.cmu.edu, jsaragih@andrew.cmu.edu
ambadar@pitt.edu

Iain Matthews
Disney Research, 4615 Forbes Ave, Pittsburgh, PA 15213
iainm@disneyresearch.com

Abstract

In 2000, the Cohn-Kanade (CK) database was released for the purpose of promoting research into automatically detecting individual facial expressions. Since then, the CK database has become one of the most widely used test-beds for algorithm development and evaluation. During this period, three limitations have become apparent: 1) While AU codes are well validated, emotion labels are not, as they refer to what was requested rather than what was actually performed, 2) The lack of a common performance metric against which to evaluate new algorithms, and 3) Standard protocols for common databases have not emerged. As a consequence, the CK database has been used for both AU and emotion detection (even though labels for the latter have not been validated), comparison with benchmark algorithms is missing, and use of random subsets of the original database makes meta-analyses difficult. To address these and other concerns, we present the Extended Cohn-Kanade (CK+) database. The number of sequences is increased by 22% and the number of subjects by 27%. The target expression for each sequence is fully FACS coded and emotion labels have been revised and validated. In addition to this, non-posed sequences for several types of smiles and their associated metadata have been added. We present baseline results using Active Appearance Models (AAMs) and a linear support vector machine (SVM) classifier using a leave-one-out subject cross-validation for both AU and emotion detection for the posed data. The emotion and AU labels, along with the extended image data and tracked landmarks will be made available July 2010.

CK数据集：
1）AU数据集经过很好的验证，但情绪标签没有，因为情绪表达的是需求而非表现。
2）缺失验证新算法的通用的能标准
3）通用数据集的标准并未制定。

1. Introduction

Automatically detecting facial expressions has become an increasingly important research area. It involves computer vision, machine learning and behavioral sciences, and can be used for many applications such as security [20], human-computer-interaction [23], driver safety [24], and health-care [17]. Significant advances have been made in the field over the past decade [21, 22, 27] with increasing interest in non-posed facial behavior in naturalistic contexts [4, 17, 25] and posed data recorded from multiple views [12, 19] and 3D imaging [26]. In most cases, several limitations are common. These include:

1. Inconsistent or absent reporting of inter-observer reliability and validity of expression metadata. Emotion labels, for instance, have referred to what expressions were requested rather than what was actually performed. Unless the validity of labels can be quantified, it is not possible to calibrate algorithm performance against manual (human) standards

2. Common performance metrics with which to evaluate new algorithms for both AU and emotion detection. Published results for established algorithms would provide an essential benchmark with which to compare performance of new algorithms.

3. Standard protocols for common databases to make possible quantitative meta-analyses.

The cumulative effect of these factors has made benchmarking various systems very difficult or impossible. This is highlighted in the use of the Cohn-Kanade (CK) database [14], which is among the most widely used datasets for

CK+数据集：数据从onset到peaks

developing and evaluating algorithms for facial expression analysis. In its current distribution, the CK (or DFAT) database contains 486 sequences across 97 subjects. Each of the sequences contains images from onset (neutral frame) to peak expression (last frame). The peak frame was reliably FACS coded for facial action units (AUs). Due to its popularity, most key recent advances in the field have evaluated their improvements on the CK database [16, 22, 25, 2, 15]. However as highlighted above, some authors employ a leave-one-out cross-validation strategy on the database, others have chosen another random train/test set configuration. Other authors have also reported results on the task of broad emotion detection even though no validated emotion labels were distributed with the dataset. The combination of these factors make it very hard to gauge the current-state-of-the-art in the field as no reliable comparisons have been made. This is a common problem across the many publicly available datasets currently available such as the MMI [19] and RUFACS [4] databases (see Zeng et al. [27] for a thorough survey of currently available databases).

In this paper, we try to address these three issues by presenting the Extended Cohn-Kanade (CK+) database, which as the name suggests is an extension to the current CK database. We have added another 107 sequences as well as another 26 subjects. The peak expression for each sequence is fully FACS coded and emotion labels have been revised and validated with reference to the FACS Investigators Guide [9] confirmed by visual inspection by emotion researchers. We propose the use of a leave-one-out subject cross-validation strategy and the area underneath the receiver operator characteristic (ROC) curve for evaluating performance in addition to an upper-bound error measure. We present baseline results on this using our Active Appearance Model (AAM)/support vector machine (SVM) system.

2. The Extended Cohn-Kanade (CK+) Dataset

2.1. Image Data

Facial behavior of 210 adults was recorded using two hardware synchronized Panasonic AG-7500 cameras. Participants were 18 to 50 years of age, 69% female, 81%, Euro-American, 13% Afro-American, and 6% other groups. Participants were instructed by an experimenter to perform a series of 23 facial displays; these included single action units and combinations of action units. Each display began and ended in a neutral face with any exceptions noted. Image sequences for frontal views and 30-degree views were digitized into either 640x490 or 640x480 pixel arrays with 8- bit gray-scale or 24-bit color values. Full details of this database are given in [14].

2.1.1 Posed Facial Expressions

In the original distribution, CK included 486 FACS-coded sequences from 97 subjects.. For the CK+ distribution, we have augmented the dataset further to include 593 sequences from 123 subjects (an additional 107 (22%) sequences and 26 (27%) subjects). The image sequence vary in duration (i.e. 10 to 60 frames) and incorporate the onset (which is also the neutral frame) to peak formation of the facial expressions

2.1.2 Non-posed Facial Expressions

During the recording of CK, 84 subjects smiled to the experimenter at one or more times between tasks. These smiles were not performed in response to a request. They comprised the initial pool for inclusion in CK+. Criteria for further inclusion were: a) relatively neutral expression at start, b) no indication of the requested directed facial action task, c) absence of facial occlusion prior to smile apex, and d) absence of image artifact (e.g., camera motion). One hundred twenty-two smiles from 66 subjects (91% female) met these criteria. Thirty two percent were accompanied by brief utterances, which was not unexpected given the social setting and hence not a criterion for exclusion.

2.2. Action Unit Labels

2.2.1 Posed Expressions

For the 593 posed sequences, full FACS coding of peak frames is provided. Approximately fifteen percent of the sequences were comparison coded by a second certified FACS coder. Inter-observer agreement was quantified with coefficient kappa, which is the proportion of agreement above what would be expected to occur by chance [10].. The mean kappas for inter-observer agreement were 0.82 for action units coded at apex and 0.75 for frame-by-frame coding. An inventory of the AUs coded in the CK+ database are given in Table 1. The FACS code coincide with the the peak frames.

2.2.2 Non-posed Expressions

A subset of action units were coded for presence/absence. These were, AU 6, AU 12, smile controls (AU 15, AU 17, AU 23/24), and AU 25/26. Comparison coding was performed for 20% of the smiles. Inter-coder agreement as measured by Cohens kappa coefficient was 0.83 for AU 6 and 0.65 for smile controls.

2.3. Validating Emotion Labels

2.3.1 Posed Expressions

We included all image data from the pool of 593 sequences that had a nominal emotion label based on the subject's impression of each of the 7 basic emotion categories: Anger,

AU	*Name*	N	AU	*Name*	N	AU	*Name*	N
1	*Inner Brow Raiser*	173	13	*Cheek Puller*	2	25	*Lips Part*	287
2	*Outer Brow Raiser*	116	14	*Dimpler*	29	26	*Jaw Drop*	48
4	*Brow Lowerer*	191	15	*Lip Corner Depressor*	89	27	*Mouth Stretch*	81
5	*Upper Lip Raiser*	102	16	*Lower Lip Depressor*	24	28	*Lip Suck*	1
6	*Cheek Raiser*	122	17	*Chin Raiser*	196	29	*Jaw Thrust*	1
7	*Lid Tightener*	119	18	*Lip Puckerer*	9	31	*Jaw Clencher*	3
9	*Nose Wrinkler*	74	20	*Lip Stretcher*	77	34	*Cheek Puff*	1
10	*Upper Lip Raiser*	21	21	*Neck Tightener*	3	38	*Nostril Dilator*	29
11	*Nasolabial Deepener*	33	23	*Lip Tightener*	59	39	*Nostril Compressor*	16
12	*Lip Corner Puller*	111	24	*Lip Pressor*	57	43	*Eyes Closed*	9

Table 1. *Frequency of the AUs coded by manual FACS coders on the CK+ database for the peak frames.*

Emotion	Criteria
Angry	AU23 and AU24 must be present in the AU combination
Disgust	Either AU9 or AU10 must be present
Fear	AU combination of AU1+2+4 must be present, unless AU5 is of intensity E then AU4 can be absent
Happy	AU12 must be present
Sadness	Either AU1+4+15 or 11 must be present. An exception is AU6+15
Surprise	Either AU1+2 or 5 must be present and the intensity of AU5 must not be stronger than B
Contempt	AU14 must be present (either unilateral or bilateral)

Table 2. *Emotion description in terms of facial action units.*

Contempt, Disgust, Fear, Happy, Sadness and Surprise. Using these labels as ground truth is highly unreliable as these impersonations often vary from the stereotypical definition outlined by FACS. This can cause error in the ground truth data which affects the training of the systems. Consequently, we have labeled the CK+ according to the FACS coded emotion labels. The selection process was in 3 steps:

1. We compared the FACS codes with the Emotion Prediction Table from the FACS maual [9]. The Emotion Prediction Table listed the facial configurations (in terms of AU combinations) of prototypic and major variants of each emotion, except contempt. If a sequence satisfied the criteria for a prototypic or major variant of an emotion, it was provisionally coded as belonging in that emotion category. In the first step, comparison with the Emotion Prediction Table was done by applying the emotion prediction "rule strictly. Applying the rule strictly means the presence of additional AU(s) not listed on the table, or the missing of an AU results in exclusion of the clip.

2. After the first pass, a more loose comparison was performed. If a sequence included an AU not included in the prototypes or variants, we determined whether they were consistent with the emotion or spoilers. For instance, AU 4 in a surprise display would be considered inconsistent with the emotion. (AU 4 is a component of negative emotion or attention and not surprise). AU 4 in the context of disgust would be considered consistent, as it is a component of negative emotion and may accompany AU 9. Similarly, we evaluated whether any necessary AU were missing. Table 2 lists the qualifying criteria. if Au were missing, Other consideration included: AU20 should not be present except for but fear; AU9 or AU10 should not be present except for disgust. Subtle AU9 or AU10 can be present in anger.

3. The third step involved perceptual judgment of whether or not the expression resembled the target emotion category. This step is not completely independent from the first two steps because expressions that included necessary components of an emotion would be likely to appear as an expression of that emotion. However, the third step was because the FACS codes only describe the expression at the peak phase and do not take into account the facial changes that lead to the peak expression. Thus, visual inspection of the clip from onset to peak was necessary to determine whether the expression is a good representation of the emotion.

As a result of this multistep selection process, 327 of the 593 sequences were found to meet criteria for one of seven discrete emotions. The inventory of this selection process is

Figure 1. Examples of the CK+ database. The images on the top level are subsumed from the original CK database and those on the bottom are representative of the extended data. All up 8 emotions and 30 AUs are present in the database. Examples of the Emotion and AU labels are: (a) Disgust - AU 1+4+15+17, (b) Happy - AU 6+12+25, (c) Surprise - AU 1+2+5+25+27, (d) Fear - AU 1+4+7+20, (e) Angry - AU 4+5+15+17, (f) Contempt - AU 14, (g) Sadness - AU 1+2+4+15+17, and (h) Neutral - AU0 are included.

Emotion	**N**
Angry (An)	45
Contempt (Co)	18
Disgust (Di)	59
Fear (Fe)	25
Happy (Ha)	69
Sadness (Sa)	28
Surprise (Su)	83

Table 3. *Frequency of the stereotypical emotion checked by manual FACS coders on the CK+ database for the peak frames.*

given in Table 3. Examples of the CK+ dataset is given in Figure 1.

2.3.2 Non-Posed Smiles

Sequences projected one at a time onto a large viewing screen to groups of 10 to 17 participants. Participants recorded their judgments during a pause following each item. They were instructed to watch the whole clip and make judgments after seeing the item number at the end of the clip. Judgments consisted of smile type (amused, embarrassed, nervous, polite, or other), and Likert-type ratings of smile intensity (from 1 = no emotion present to 7 = extreme emotion), and confidence in smile type judgments (from 1 = no confidence to 7 = extreme confidence).

For each sequence we calculated the percentage of participants who judged it as amused, embarrassed, nervous, polite, or other. These percentages are referred to as judgment scores. From the five judgment scores, smiles were assigned to the modal type if at least 50% of participants endorsed that type and no more than 25 % endorsed another. The 50% endorsement criterion represented the minimum modal response. The 25% maximum endorsement for the rival type was used to ensure discreteness of the modal response. By this criterion, 19 were classified as perceived amused, 23 as perceived polite, 11 as perceived embarrassed or nervous, and 1 as other. CK+ includes the modal scores and the ratings for each sequence. For details and for future work using this portion of the database please see and cite [1].

3. Baseline System

In our system, we employ an Active Appearance Model (AAM) based system which uses AAMs to track the face and extract visual features. We then use support vector machines (SVMs) to classify the facial expressions and emotions. An overview of our system is given in Figure 2. We describe 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 [7].

The shape $\mathbf{s}$ of an AAM [7] is described by a 2D triangulated mesh. In particular, the coordinates of the mesh vertices define the shape $s = [x_1, y_1, x_2, y_2, \ldots, x_n, y_n]$, where n is the number of vertices. These vertex locations correspond to a source appearance image, from which the shape was aligned. Since AAMs allow linear shape variation, the shape s can be expressed as a base shape $\mathbf{s}_0$ plus a

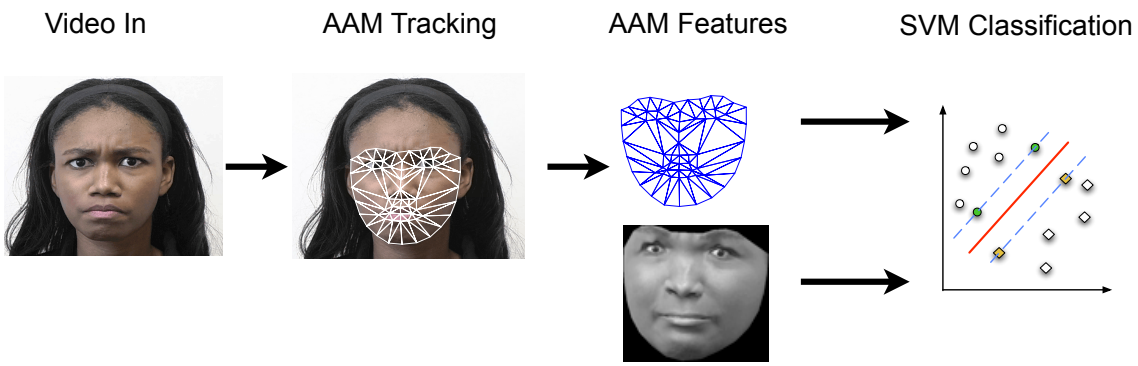

Figure 2. Block diagram of our automatic system. The face is tracked using an AAM and from this we get both similarity-normalized shape (SPTS) and canonical appearance (CAPP) features. Both these features are used for classification using a linear SVM.

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 $\mathbf{p} = (p_1, \ldots, p_m)^T$ are the shape parameters. These shape parameters can typically be divided into rigid similarity parameters $\mathbf{p}_s$ and non-rigid object deformation parameters $\mathbf{p}_o$, such that $\mathbf{p}^T = [\mathbf{p}_s^T, \mathbf{p}_o^T]$. Similarity parameters are associated with a geometric similarity transform (i.e. translation, rotation and scale). The object-specific parameters, are the residual parameters representing non-rigid geometric variations associated with the determing object shape (e.g., mouth opening, eyes shutting, etc.). Procrustes alignment [7] 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 fitting algorithm described in [18].

3.2. Feature Extraction

Once we have tracked the patient's face by estimating the shape and appearance AAM parameters, we can use this information to derive the following features:

- **SPTS:** The similarity normalized shape, $\mathbf{s}_n$, refers to the 68 vertex points in 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 similarity normalized shape features, SPTS, is given in Figure 2. AU0 normalization was used in this work, by subtracting the features of the first frame (which was neutral).
- **CAPP:** The 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. For this study, the resulting 87×93 synthesized grayscale image was used. In previous work [3], 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.

3.3. Support Vector Machine Classification

Support vector machines (SVMs) have been proven useful in a number of pattern recognition tasks including face and facial action recognition. 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,

$$\mathbf{w}^T \mathbf{x}^* \quad \begin{matrix} >^{true} \\ <_{false} \end{matrix} \quad b \tag{2}$$

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. A linear kernel was used in our experiments due to its ability to generalize well to unseen data in many pattern recognition tasks [13]. LIBSVM was used for the training and testing of SVMs [6].

For AU detection, we just used a linear one-vs-all two-class SVM (i.e. AU of interest vs non-AU of interest). For the training of the linear SVM for each of the AU detectors,

AU	N	SPTS	CAPP	SPTS+CAPP
1	173	**94.1** ± 1.8	**91.3** ± 2.1	**96.9** ± 1.3
2	116	**97.1** ± 1.5	**95.6** ± 1.9	**97.9** ± 1.3
4	191	**85.9** ± 2.5	**83.5** ± 2.7	**91.0** ± 2.1
5	102	**95.1** ± 2.1	**96.6** ± 1.8	**97.8** ± 1.5
6	122	**91.7** ± 2.5	**94.0** ± 2.2	**95.8** ± 1.8
7	119	**78.4** ± 3.8	**85.8** ± 3.2	**89.2** ± 2.9
9	74	**97.7** ± 1.7	**99.3** ± 1.0	**99.6** ± 0.7
11	33	**72.5** ± 7.8	**82.0** ± 6.7	**85.2** ± 6.2
12	111	**91.0** ± 2.7	**96.0** ± 1.9	**96.3** ± 1.8
15	89	**79.6** ± 4.3	**88.3** ± 3.4	**89.9** ± 3.2
17	196	**84.4** ± 2.6	**90.4** ± 2.1	**93.3** ± 1.8
20	77	**91.0** ± 3.3	**93.0** ± 2.9	**94.7** ± 2.6
23	59	**91.1** ± 3.7	**87.6** ± 4.3	**92.2** ± 3.5
24	57	**83.3** ± 4.9	**90.4** ± 3.9	**91.3** ± 3.7
25	287	**97.1** ± 1.0	**94.0** ± 1.4	**97.5** ± 0.9
26	48	**75.0** ± 6.3	**77.6** ± 6.0	**80.3** ± 5.7
27	81	**99.7** ± 0.7	**98.6** ± 1.3	**99.8** ± 0.5
AVG		**90.0** ± 2.5	**91.4** ± 2.4	**94.5** ± 2.0

Table 4. *Results showing the area underneath the ROC curve for the shape and appearance features for AU detection. Note the average is a weighted one, depending on the number of positive examples.*

all neutral and peak frames from the training sets were used. The frames which were coded to contain the AU were used as positive examples and all others were used as negative examples, regardless if the AU occurred alone or in combination with other AUs. The output from SVM related just to the distance to the hyperplane which works well for a single decision. However, these scores have no real meaning when comparing them from different SVMs. As such, comparing or combining these scores does not make sense and can lead to erroneous results. Calibrating the scores into a common domain is required so that comparisons and fusion can take place. Logistical linear regression is one method of doing this [5]. In this paper, we fused both the scores from the SPTS and CAPP feature sets to determine if there was any complementary information between these two. The FoCal package was used for calibrating and fusing the various AU SVM scores together using LLR [5].

For the task of emotion detection a forced multi-class decision has to be made. To accommodate this, a one-versus all multi-class SVM was used (i.e. Angry vs not Angry, Happy vs not Happy etc.). All frames which were coded as the particular emotion of interest were used as the positive example and all others were used as negative examples. A seven-way forced choice of the possible emotions was performed (neutral was neglected as the neutral frame was subtracted from all features).

4. Experiments

4.1. Benchmarking Protocol and Evaluation Metric

In this paper, we document two types of experiments that can be conducted on the posed section of the CK+ database: (i) AU detection, and (ii) emotion detection. To maximize the amount of training and testing data, we believe the use of a leave-one-subject-out cross-validation configuration should be used. This means for AU detection, 123 different training and testing sets need to be used, and for emotion detection, 118 different training and test sets need to be used.

In terms of evaluating the different experiments, for AU detection the area underneath the receiver-operator characteristic (ROC) curve is a reliable measure. This curve is obtained by plotting the hit-rate (true positives) against the false alarm rate (false positives) as the decision threshold varies. The area underneath this curve (A'), is used to assess the performance [4]. The A' metric ranges from 50 (pure chance) to 100 (ideal classification)[1]. Results should be averaged across these sets. An upper-bound on the uncertainty of the A' statistic should also be included to give an idea of the reliability of the performance. A common statistic used is for this is $s = \sqrt{\frac{A'(1-A')}{\min\{n_p, n_n\}}}$ where n_p, n_n are the number of positive and negative examples [8, 25]. For emotion detection, we used a confusion matrix to document the results.

4.2. AU Detection Results

The results for AU detection for both similarity-normalized shape (SPTS) and canonical appearance (CAPP) features and the combination of both features are given in Table 4. From the results it can be seen that all feature types achieve very good overall accuracy with performance of $A' >= 90$, with a combination of both the SPTS+CAPP features yielding the best performance with a rating of 94.5. This suggests that there exists complimentary information between both shape and appearance features.

In terms of individual AU detection, it can be seen depending on the AU, the best performing feature set varies. When comparing the individual SPTS and CAPP features, the SPTS features yielded the higher detection rates for AUs 1, 2, 4, 23, 25 and 27, while the CAPP features gained better performance for AUs 5, 6, 7, 9, 11, 12, 15, 17, 20, 24 and 26. Even though the difference in performance is quite small for some of these AUs, an explanation of these results can stem from the AAM 2-D mesh. For example AUs 1, 2 and 4 are actions coinciding with eye brow movement which can easily be picked up by the shape features as they

[1] In 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

lie on the AAM 2-D mesh. For AUs 6, 9 and 11, a lot of textural change in terms of wrinkles and not so much in terms of contour movement, which would suggest why the CAPP features performed better than the SPTS for these.

4.3. Emotion Detection Results

The performance of the shape (SPTS) features for emotion detection is given in Table 5 and it can be seen that Disgust, Happiness and Surprise all perform well compared to the other emotions. This result is intuitive as these are very distinctive emotions causing a lot of deformation within the face. The AUs associated with these emotions also lie on the AAM mesh, so movement of these areas is easily detected by our system. Conversely, other emotions (i.e. Anger, Sadness and Fear) that do not lie on the AAM mesh do not perform as well for the same reason. However, for these emotions textural information seems to be more important. This is highlighted in Table 6. Disgust also improves, as there is a lot of texture information contained in the nose wrinkling (AU9) associated with this prototypic emotion.

For both the SPTS and CAPP features, Contempt has a very low hit-rate. However, when output from both the SPTS and CAPP SVMs are combined (through summing the output probabilities) it can be seen that the detection of this emotion jumps from just over 20% to over 80% as can be seen in Table 7. An explanation for this can be from that fact that this emotion is quite subtle and it gets easily confused with other, stronger emotions. However, the confusion does not exist in both features sets. This also appears to have happen for the other subtler emotions such as Fear and Sadness, with both these emotions benefitting from the fusion of both shape and appearance features.

The results given in Table 7 seem to be inline with recent perceptual studies. In a validation study conducted on the Karolinska Directed Emotional Faces (KDEF) database [11], results for the 6 basic emotions (i.e. all emotions in CK+ except Contempt) plus neutral, were similar to the ones presented here. In this study they used 490 images (i.e. 70 per emotion) and the hit rates for each emotion were[2]: Angry - 78.81% (75.00%), Disgust - 72.17% (94.74%), Fear - 43.03% (65.22%), Happy - 92.65% (100%), Sadness - 76.70% (68.00%), Surprised - 96.00% (77.09%), Neutral - 62.64% (100%)[3].

This suggests that an automated system can do just as a good job, if not better as a naive human observer and suffer from the same confusions due to the perceived ambiguity between subtle emotions. However, human observer ratings need to be performed on the CK+ database and automated results need to be conducted on the KDEF database to test out the validity of these claims.

[2]our results are in the brackets next to the KDEF human observer results

[3]As neutral frame is subtracted, just a simple energy based measure of shape movement results in perfect detection of the neutral frame. However, this is very unlikely to occur in a realistic scenario as there is much variation in the neutral frame and subtracted this will not be possible

	An	Di	Fe	Ha	Sa	Su	Co
An	**35.0**	40.0	0.0	5.0	5.0	15.0	0.0
Di	7.9	**68.4**	0.0	15.8	5.3	0.0	2.6
Fe	8.7	0.0	**21.7**	21.7	8.7	26.1	13.0
Ha	0.0	0.0	0.0	**98.4**	1.6	0.0	0.0
Sa	28.0	4.0	12.0	0.0	4.0	**28.0**	24.0
Su	0.0	0.0	0.0	0.0	0.0	**100.0**	0.0
Co	15.6	3.1	6.3	0.0	15.6	34.4	**25.0**

Table 5. *Confusion matrix of emotion detection for the similarity-normalized shape (SPTS) features - the emotion classified with maximum probability was shown to be the emotion detected.*

	An	Di	Fe	Ha	Sa	Su	Co
An	**70.0**	5.0	5.0	0.0	10.0	5.0	5.0
Di	5.3	**94.7**	0.0	0.0	0.0	0.0	0.0
Fe	8.7	0.0	**21.7**	21.7	8.7	26.1	13.0
Ha	0.0	0.0	0.0	**100.0**	0.0	0.0	0.0
Sa	16.0	4.0	8.0	0.0	**60.0**	4.0	8.0
Su	0.0	0.0	1.3	0.0	0.0	**98.7**	0.0
Co	12.5	12.5	3.1	0.0	28.1	21.9	**21.9**

Table 6. *Confusion matrix of emotion detection for the canonical appearance (CAPP) features - the emotion classified with maximum probability was shown to be the emotion detected.*

	An	Di	Fe	Ha	Sa	Su	Co
An	**75.0**	7.5	5.0	0.0	5.0	2.5	5.0
Di	5.3	**94.7**	0.0	0.0	0.0	0.0	0.0
Fe	4.4	0.0	**65.2**	8.7	0.0	13.0	8.7
Ha	0.0	0.0	0.0	**100.0**	0.0	0.0	0.0
Sa	12.0	4.0	4.0	0.0	**68.0**	4.0	8.0
Su	0.0	0.0	0.0	0.0	4.0	**96.0**	0.0
Co	3.1	3.1	0.0	6.3	3.1	0.0	**84.4**

Table 7. *Confusion matrix of emotion detection for the combination of features (SPTS+CAPP) - the emotion classified with maximum probability was shown to be the emotion detected. The fusion of both systems were performed by summing up the probabilities from output of the multi-class SVM.*

5. Conclusions and Future Work

In this paper, we have described the Extended Cohn-Kanade (CK+) database for those researchers wanting to prototype and benchmark systems for automatic facial expression detection. Due to the popularity and ease of access for the original Cohn-Kanade dataset this is seen as a very valuable addition to the already existing corpora that

is already in existence. For a fully automatic system to be robust for all expression in a myriad of realistic scenarios more data is required. For this to occur very large reliably coded datasets across a wide array of visual variabilities are required (at least 5 to 10k examples for each action). This will require a concerted collaborative research effort from a wide array of research institutions due to the cost associated with capturing, coding, storing and distributing such data. In this final distribution of the database, we hope to augment what we described here with non-frontal data consisting of synchronous views of the posed expressions from an angle of 30 degrees.

6. CK+ Database Availability

In the summer ($\approx$ July) of 2010 we envisage that the CK+ database will be ready for distribution to the research community. Similarly to the original CK database, interested parties will have to visit http://vasc.ri.cmu.edu/idb/html/face/facial_expression/ and download and sign an agreement that governs its use and return the completed form. Once the form has been received, it will take around 4 to 5 business days for receipt of instructions by email on how to download the database. For work being conducted on the non-posed data, please cite [1].

7. Acknowledgements

This project was supported in part by National Institute of Mental Health grant R01 MH51435. Special mention also needs to go to Nicole Ridgeway for providing technical assistance.

References

[1] Z. Ambadar, J. Cohn, and L. Reed. All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior*, 33:17–34, 2009. 4, 8

[2] A. Ashraf, S. Lucey, and T. Chen. Re-Interpreting the Application of Gabor Filters as a Manipulation of the Margin in Linear Support Vector Machines. *IEEE Transactions on Pattern Analysis and Machine Intelligence (in press)*, 2010. 2

[3] 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, Nagoya, Aichi, Japan, 2007. ACM. 5

[4] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Automatic Recognition of Facial Actions in Spontaneous Expressions. *Journal of Multimedia*, 2006. 1, 2, 6

[5] N. Brummer and J. du Preez. Application-Independent Evaluation of Speaker Detection. *Computer Speech and Language*, 2005. 6

[6] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm. 5

[7] T. Cootes, G. Edwards, and C. Taylor. Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001. 4, 5

[8] C. Cortes and M. Mohri. Confidence Intervals for the Area Under the ROC curve. *Advances in Neural Information Processing Systems*, 2004. 6

[9] P. Ekman, W. Friesen, and J. Hager. *Facial Action Coding System: Research Nexus*. Network Research Information, Salt Lake City, UT, USA, 2002. 2, 3

[10] J. Fleiss. *Statistical Methods for Rates and Proportions*. Wiley, N.Y, 1981. 2

[11] E. Goeleven, R. D. Raedt, L. Leyman, and B. Verschuere. The Karolinska Directed Emotional Faces: A validation study. *Cognition and Emotion*, 22(6):1094–1118. 7

[12] R. Gross, I. Matthews, S. Baker, and T. Kanade. The CMU Multiple Pose, Illumination, and Expression (MultiPIE). Technical report, Robotics Institute, Carnegie Mellon University, 2007. 1

[13] C. Hsu, C. C. Chang, and C. J. Lin. A practical guide to support vector classification. Technical report, 2005. 5

[14] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 46–53, 2000. 1, 2

[15] S. Koelstra, M. Pantic, and I. Patras. A Dynamic Texture Based Approach to Recognition of Facial Actions and Their Temporal Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (in press)*, 2010. 2

[16] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of Facial Expression Extracted Automatically from Video. *Journal of Image and Vison Computing*, 24(6):615–625, 2006. 2

[17] P. Lucey, J. Cohn, S. Lucey, I. Matthews, S. Sridharan, and K. Prkachin. Automatically Detecting Pain Using Facial Actions. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, pages 1–8, 2009. 1

[18] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004. 5

[19] M. Pantic, M. Valstar, R. Rademaker, and L. Maat. Fully automatic Facial Action Recognition in Spontaneous Behavior. In *Proceedings of the International Conference on Multimedia and Expo*, pages 317–321, 2005. 1, 2

[20] A. Ryan, J. Cohn, S. Lucey, J. Saragih, P. Lucey, F. D. la Torre, and A. Rossi. Automated Facial Expression Recognition System. In *Proceedings of the International Carnahan Conference on Security Technology*, pages 172–177, 2009. 1

[21] 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. 1

[22] 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. 1, 2

[23] A. Vinciarelli, M. Pantic, and H. Bourlard. Social Signal Processing: Survey of an Emerging Domain. *Image and Vision Computing*, 31(1):1743–1759, 2009. 1

[24] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan. Automated Drowsiness Detection for Improved Driver Saftey-Comprehensive Databases for Facial Expression Analysis. In *Proceedings of the International Conference on Automotive Technologies*, 2008. 1

[25] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan. Towards Practical Smile Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):2106–2111, 2009. 1, 2, 6

[26] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A high-resolution 3D dynamic facial expression database. In *Proceedings of the International Conference on Automatic Face And Gesture Recognition*, 2008. 1

[27] Z. Zeng, M. Pantic, G. Roisman, and T. Huang. A Survey of Affect Recognition Methods: Audio, Visual and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009. 1, 2