AUTOMATIC DETECTION AND INTENSITY ESTIMATION OF SPONTANEOUS SMILES

by

Jeffrey M. Girard

B.A. in Psychology/Philosophy, University of Washington, 2009

Submitted to the Graduate Faculty of

The Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

THE DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Jeffrey M. Girard

It was defended on

June 7th 2013

and approved by

J. F. Cohn, Ph.D., Professor

M. A. Sayette, Ph.D., Professor

F. De la Torre, Ph.D., Professor

K. A. Roecklein, Ph.D., Professor

Thesis Advisor: J. F. Cohn, Ph.D., Professor

AUTOMATIC DETECTION AND INTENSITY ESTIMATION OF SPONTANEOUS SMILES

Jeffrey M. Girard, M.S.

University of Pittsburgh, 2013

Both the occurrence and intensity of facial expression are critical to what the face reveals. While much progress has been made towards the automatic detection of expression occurrence, controversy exists about how best to estimate expression intensity. Broadly, one approach is to adapt classifiers trained on binary ground truth to estimate expression intensity. An alternative approach is to explicitly train classifiers for the estimation of expression intensity. We investigated this issue by comparing multiple methods for binary smile detection and smile intensity estimation using two large databases of spontaneous expressions. SIFT and Gabor were used for feature extraction; Laplacian Eigenmap and PCA were used for dimensionality reduction; and binary SVM margins, multiclass SVMs, and ε-SVR models were used for prediction. Both multiclass SVMs and ε-SVR classifiers explicitly trained on intensity ground truth outperformed binary SVM margins for smile intensity estimation. A surprising finding was that multiclass SVMs also outperformed binary SVM margins on binary smile detection. This suggests that training on intensity ground truth is worthwhile even for binary expression detection.

TABLE OF CONTENTS

1.0 INTRODUCTION 1

1.1 Previous Work 3

1.2 The Current Study 4

2.0 METHODS 6

2.1 Participants and Data 6

2.1.1 BP4D Database 6

2.1.2 Spectrum Database 6

2.2 Manual Expression Annotation 7

2.2.1 AU Occurrence 7

2.2.2 AU Intensity 8

2.3 Automatic Expression Annotation 9

2.3.1 Tracking 9

2.3.2 Extraction 9

2.3.3 Reduction 10

2.3.4 Cross-validation 11

2.3.5 Prediction 12

2.4 Performance Evaluation 13

2.5 Data Analysis 14

3.0 RESULTS 15

3.1 Smile Intensity Estimation 15

3.2 Binary Smile Detection 19

4.0 DISCUSSION 20

4.1 Smile Intensity Estimation 20
4.2 Binary Smile Detection 21
4.3 Conclusions 22
4.4 Limitations and Future Directions 23
5.0 ACKNOWLEDGMENTS 24
BIBLIOGRAPHY 25

LIST OF TABLES

1 General Linear Model results for Smile Intensity Estimation (ICC) 17
2 General Linear Model Results for Binary Smile Detection (F_1) 18

LIST OF FIGURES

1 Flowchart of Automatic Expression Annotation Methods 7
2 Smile Intensity Levels 8
3 Visualization of Tracking Procedures 10
4 Visualization of Extraction Procedures 11
5 Smile Intensity Estimation Performance 16
6 Binary Smile Detection Performance 16

1.0 INTRODUCTION

The face is an important avenue of communication capable of regulating social interaction and providing the careful observer with a wealth of information. Facial expression analysis has informed psychological studies of emotion [15, 22, 84], intention [30, 52], physical pain [56, 70], and psychopathology [12, 33], among other topics. It is also central to computer science research on human-computer interaction [14, 69] and computer animation [68].

There are two general approaches to classifying facial expression. Message-based approaches seek to identify the meaning of each expression; this often takes the form of classifying expressions into one of six basic emotions: anger, disgust, fear, happiness, sadness, or surprise [45]. However, this involves a great deal of interpretation and fails to account for the fact that facial expressions serve a communicative function [30], can be controlled or dissembled [20], and often depend on context for interpretation [3]. Sign-based approaches, on the other hand, describe changes in the face during an expression rather than attempting to capture its meaning. By separating description from interpretation, sign-based approaches achieve more objectivity and comprehensiveness.

The most commonly used sign-based approach for describing facial expression is the Facial Action Coding System (FACS) [24], which decomposes facial expressions into component parts called action units. Action units (AU) are anatomically-based and correspond to the contraction of specific facial muscles. AU may occur alone or in combination with others to form complex facial expressions. They may also vary in intensity (i.e., magnitude of muscle contraction). The FACS manual provides coders with detailed descriptions of the shape and appearance changes necessary to identify each AU and its intensity.

Much research using FACS has focused on the occurrence and AU composition of different expressions [21]. For example, smiles including the orbicularis oculi muscle (i.e., AU6)

are more likely to occur during pleasant circumstances [23, 28] and smiles including the buccinator muscle (i.e., AU14) are more likely to occur during active depression [71, 33]. However, a promising subset of research has begun to focus on what can be learned about and from the intensity of expressions. This work has shown that expression intensity is linked to both the intensity of emotional experience and the sociality of the context [22, 29, 38]. For example, Hess et al. [38] found that participants were the most expressive when experiencing strong emotions in the company of friends. Other studies have used the intensity of facial expressions (e.g., in yearbook photos) to predict a number of social and health outcomes years later. For example, smile intensity in a posed photograph has been linked to later life satisfaction, marital status (i.e., likelihood of divorce), and even years lived [1, 36, 37, 67, 74].

It is likely that research has only begun to scratch the surface of what might be learned from expressions' intensities. Intensity estimation is also critical to the modeling of an expression's temporal dynamics (i.e., changes in intensity over time). Temporal dynamics is a relatively new area of study, but has already been linked to expression interpretation, person perception, and psychopathology. For example, the speed with which a smile onsets and offsets has been linked to interpretations of the expression's meaning and authenticity, as well as to ratings of the smiling person's attractiveness and personality [54]. Expression dynamics have also been used to explore the nonverbal behavior of individuals with depression, schizophrenia, and obsessive-compulsive disorder [63, 64, 49].

However, the manual coding of AU occurrence is already highly time-consuming without coding for frame-level intensity. Manually coding one minute of video for AU onsets and offsets can take anywhere between one and five hours. To make this kind of coding more manageable, there has been a great deal of research interest in the automatic detection of facial expressions [e.g., 25, 84]. However, the vast majority of this work has focused on the binary detection of expressions (i.e., predicting their presence or absence). Fewer studies have tested different methods for the automatic estimation of facial expression intensity.

In an early and influential work on this topic, Bartlett et al. [4] applied standard binary expression detection techniques to estimate expressions' peak intensity. This and subsequent work [5, 6] encouraged the use of the margins and posterior probabilities of binary-trained classifiers as proxies for expression intensity. The assumption underlying this practice is

that the classifier's margin will be positively correlated with the expression's intensity, i.e., the classifier will be more confident about more intense expressions. Yang et al. [83] and others have challenged this assumption, arguing that there is no logical necessity for such a correlation and that factors other than an expression's intensity (e.g., similarity to the training set, concurrent facial movements, and image artifacts) are likely to influence a classifier's confidence. However, relatively few studies have investigated this question empirically. Are binary expression detection and expression intensity estimation fundamentally different problems that require different solutions, as Yang et al. [83] assert, or can the same methods and even the same classifiers be used for both, as Bartlett et al. [4] have suggested?

1.1 PREVIOUS WORK

Since Bartlett et al. [4], many studies have used classifier margins and posterior probabilities to estimate expression intensity [5, 6, 55, 72, 53, 81, 83, 73, 75, 76]. However, only a few of them have quantitatively evaluated their performance by comparing their estimations to manual (i.e., "ground truth") coding. Several studies [6, 81, 73] found that classifier margin and expression intensity were positively correlated during posed expressions. However, such correlations have typically been lower during spontaneous expressions. In a highly relevant study, Whitehill et al. [81] focused on the estimation of spontaneous smile intensity and found a high correlation between classifier margin and smile intensity. However, this was in five short video clips and it is unclear how the ground truth intensity coding was obtained.

Recent studies have also used methods other than classifier margins for intensity estimation, such as regression [50, 73, 18, 51, 47] and multiclass classifiers [60, 66, 62]. These studies have found that the predictions of support vector regression models and multiclass classifiers were typically highly correlated with expression intensity during both posed and spontaneous expressions. Finally, several studies [11, 17, 65] used extracted features to estimate expression intensity directly. For example, Messinger et al. [65] found that mouth radius was highly correlated with spontaneous smile intensity in five video clips.

Very few studies have compared different estimation methods using the same data and

performance evaluation methods. Savran et al. [73] found that support vector regression outperformed the margins of binary support vector machine classifiers on the intensity estimation of posed expressions. Ka Keung and Yangsheng [50] found that support vector regression outperformed cascading neural networks on the intensity estimation of posed expressions, and Dhall and Goecke [18] found that Gaussian process regression outperformed both kernel partial least squares and support vector regression on the intensity estimation of posed expressions. Yang et al. [83] also compared classifier margins with an intensity-trained model, but used their outputs to rank images by intensity rather than to estimate it.

Much of the previous work has been limited in three ways. First, many studies [17, 18, 50, 83] adopted a message-based approach, which is problematic for the reasons described earlier. Second, the majority of this work [17, 18, 50, 73, 83] focused on posed expressions, which limits the external validity and generalizability of their findings. Third, most of these studies were limited in terms of the ground truth they compared their estimations to. Some studies [4, 5, 6] only coded expressions' peak intensities, while others [60, 65, 66, 81] obtained frame-level ground truth, but only for a handful of subjects. Without a large amount of expert-coded, frame-level ground truth, it is impossible to truly gauge the success of an automatic intensity estimation system.

1.2 THE CURRENT STUDY

The main contribution of the current study was to test the hypothesis that binary expression detection and expression intensity estimation are fundamentally different problems that require different solutions. We explored this hypothesis by comparing different techniques using the same data and performance evaluation methods. We also improve upon previous work by using a sign-based approach, two large datasets of spontaneous expressions, and expert-coded, frame-level ground truth.

Techniques for feature extraction, dimensionality reduction, and classification/regression were systematically varied, enabling us to see whether the methods that were best for binary detection were also best for intensity estimation. Based on the findings of previous studies,

we hypothesized that methods trained on intensity ground truth would outperform methods trained on binary ground truth. We also tested whether binary-trained models could be applied to estimate expression intensity and whether intensity-trained models could be applied to detect expression occurrence. Smiles were chosen for this in-depth analysis because they are the most commonly occurring facial expression [7], are implicated in affective displays and social signaling [39, 40], and appear in much of the previous work on both automatic intensity estimation and the psychological exploration of facial expression intensity.

2.0 METHODS

2.1 PARTICIPANTS AND DATA

In order to increase the sample size and explore the generalizability of the findings, data was drawn from two separate datasets. Both datasets recorded and FACS coded participant facial behavior during a non-scripted, spontaneous dyadic interaction. They differ in terms of the context of the interaction, the demographic makeup of the sample, and the constraints placed upon data collection (e.g., illumination, frontality, and head motion). Because of how its segments were selected, the BP4D database also had more frequent and intense smiles.

2.1.1 BP4D Database

FACS coded video was available for 30 adults (50% female, 50% white, mean age 20.7 years) from the Binghamton-Pittsburgh 4D (BP4D) Database [85]. Participants were filmed with both a 3D dynamic face capturing system and a 2D frontal camera (520x720 pixel resolution) while engaging in eight tasks designed to elicit emotions such as anxiety, surprise, happiness, embarrassment, fear, pain, anger, and disgust. Facial behavior from the 20-second segment with the highest AU density (i.e., frequency and intensity) from each task was coded from the 2D video. The BP4D database will be publicly available in the summer of 2013.

2.1.2 Spectrum Database

FACS coded video was available for 33 adults (67.6% female, 88.2% white, mean age 41.6 years) from the Spectrum database [12]. The participants suffered from major depressive disorder [2] and were recorded during clinical interviews to assess symptom severity over

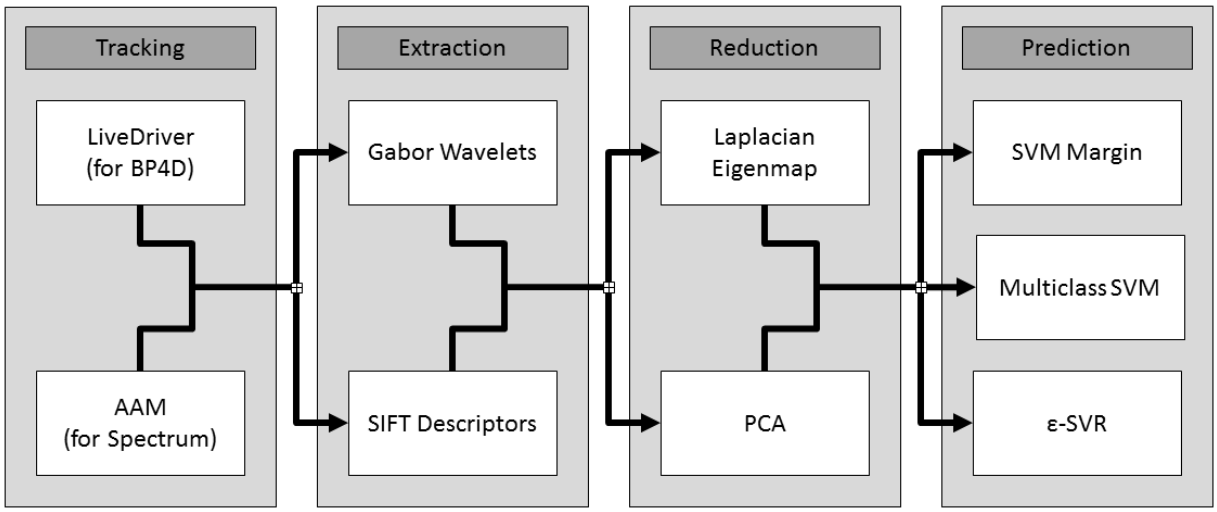

Figure 1: Flowchart of Automatic Expression Annotation Methods

the course of treatment [34]. A total of 69 interviews were recorded using four hardware-synchronized analogue cameras. Video from a camera roughly 15 degrees to the participant's right was digitized into 640x480 pixel arrays for analysis. Facial behavior during the first three interview questions (about depressed mood, feelings of guilt, and suicidal ideation) was coded; these segments ranged in length from 28 to 242 seconds (M = 100 seconds). The Spectrum database is not publicly available.

2.2 MANUAL EXPRESSION ANNOTATION

2.2.1 AU Occurrence

For the BP4D database, participant facial behavior was manually FACS coded from video by certified coders. Event onset and offset were coded for 34 commonly occurring AU. Inter-observer reliability for AU12 occurrence was F_1=0.96. For the Spectrum database, participant facial behavior was manually FACS coded from video by certified coders. Event

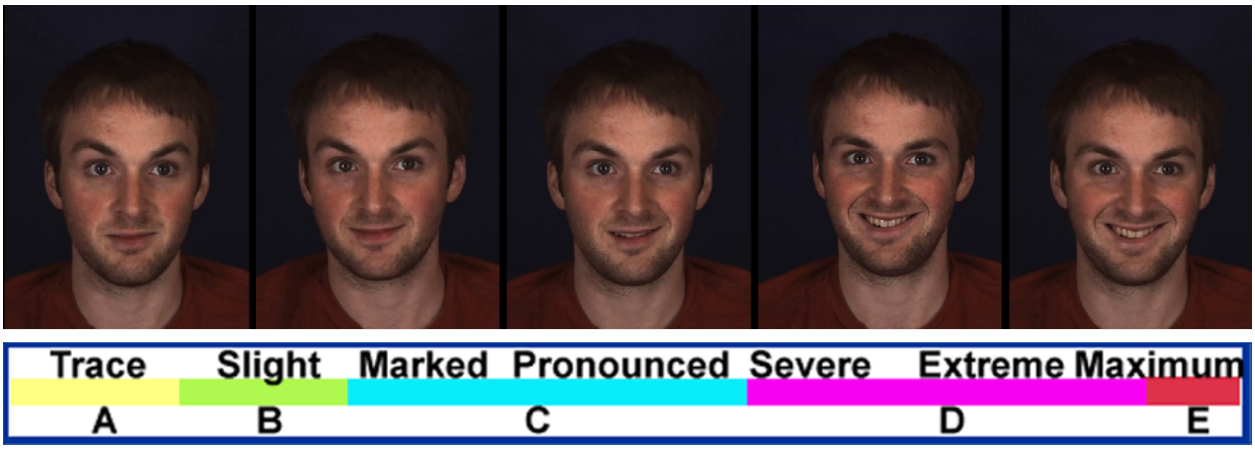

Figure 2: Smile (i.e., AU12) intensity levels defined by the FACS manual [19].

onset, offset, and apex were coded for 17 commonly occurring AU. Inter-observer reliability for AU12 occurrence was F_1=0.71. For both datasets, onsets and offsets were converted to frame-level codes with 0 and 1 representing the absence and presence of AU12, respectively.

2.2.2 AU Intensity

The manual FACS coding procedures described in subsection 2.2.1 were used to identify the onsets and offsets of AU12 events. Separate video clips of each event were generated and coded for intensity by certified coders using continuous measurement software [32]. This coding involved assigning each video frame an integer value between 0 and 5, with 0 representing the absence of AU12 and 1 through 5 representing trace through maximum intensity (Figure 2) as defined by the FACS manual [24]. Ten percent of clips were independently coded by a second certified FACS coder to establish reliability, which was ICC=0.92.

2.3 AUTOMATIC EXPRESSION ANNOTATION

2.3.1 Tracking

Facial landmark points indicate the location of important facial components (e.g., eye corners, nose tip, lip corners). For the BP4D database, sixty-four facial landmarks (Figure 3) were tracked in each video frame using the LiveDriver SDK from Image Metrics [44]. Overall, 4% of video frames were untrackable, mostly due to occlusion or extreme out-of-plane rotation. A global normalizing (i.e., similarity) transformation was applied to the data for each video frame to remove variation due to rigid head motion. Finally, each image was cropped to the area surrounding the detected face and scaled to 128x128 pixels.

For the Spectrum database, sixty-six facial landmarks (Figure 3) were tracked using active appearance models (AAM) [13]. AAM is a powerful approach that combines the shape and texture variation of an image into a single statistical model. Approximately 3% of video frames were manually annotated for each subject and then used to build the AAMs. The frames then were automatically aligned using a gradient-descent AAM fitting algorithm [61]. Overall, 9% of frames were untrackable, again mostly due to occlusion and rotation. The same normalization procedures used on the LiveDriver landmarks were also used on the AAM landmarks. Additionally, because AAM includes landmark points along the jawline, we were able to remove non-face information from the images using a convex hull algorithm.

2.3.2 Extraction

Two types of appearance features were extracted from the tracked and normalized faces. Following previous work on expression detection [10] and intensity estimation [4, 73], Gabor wavelets [16, 27] were extracted in localized regions surrounding each facial landmark point (left Figure 4). Gabor wavelets are biologically-inspired filters, operating in a similar fashion to simple receptive fields in mammalian visual systems [48]. They have been found to be robust to misalignment and changes in illumination [57]. By applying a filter bank of eight equally-spaced orientations and five scales (i.e., 17, 23, 33, 46, 65 pixels) at each localized region, specific changes in facial texture and orientation (which map onto facial wrinkles,

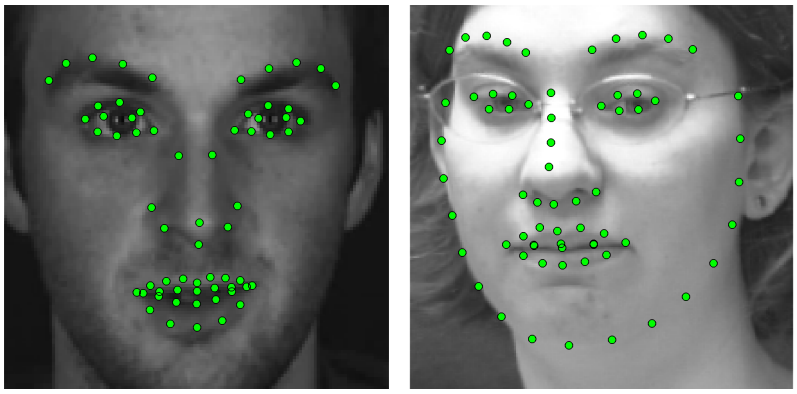

Figure 3: Visualization of Tracking Procedures. Left: LiveDriver landmarks in BP4D [85]. Right: AAM landmarks in Spectrum [12].

folds, and bulges) were quantified. Scale-invariant feature transforms (SIFT) [59, 80] were also extracted in localized regions surrounding each facial landmark point (Figure 4). As its name implies, SIFT is invariant to image scaling, translation, and rotation, and partially invariant to illumination changes and affine or 3D projection. By applying a geometric descriptor (scale=3, orientation=0) to each facial landmark, changes in facial texture and orientation were quantified.

2.3.3 Reduction

Both types of features exhibited high dimensionality, which makes classification a difficult and resource-intensive problem. Two approaches for dimensionality reduction were compared on their ability to yield discriminant features for classification. Principal Components Analysis (PCA) [82] is a linear technique used to project a feature vector from a high dimensional space into a low dimensional space. Unsupervised PCA was used to find the smallest number of dimensions that accounted for 95% of the variance in a randomly selected sample of 100,000 frames. This technique reduced the Gabor features from 2640 dimensions per video frame to 162, and reduced the SIFT features from 8448 dimensions per video frame to 362. Laplacian Eigenmap [8] is a nonlinear technique used to learn the low dimensional

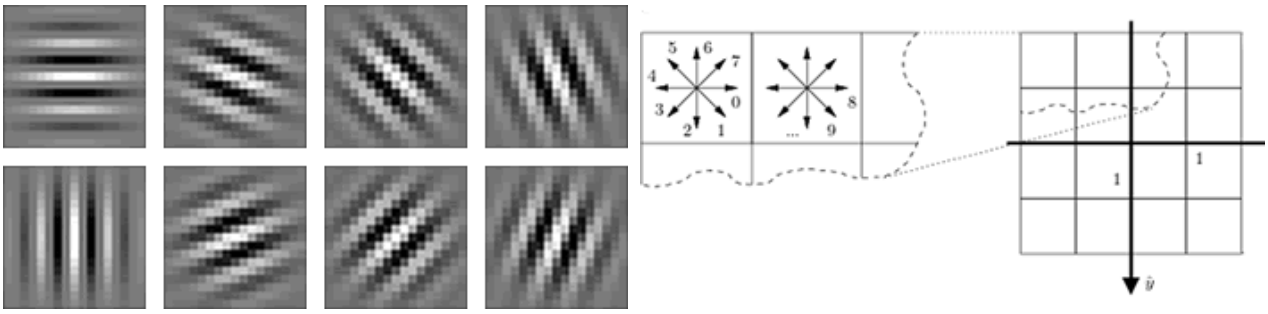

Figure 4: Visualization of Extraction Procedures. Left: Eight orientations of Gabor wavelets. Right: Geometric indexing of the SIFT descriptor.

manifold that the original (i.e., high dimensional) feature data lies upon. Following recent work by Mahoor et al. [60], supervised Laplacian Eigenmaps were trained on a randomly selected sample of 2500 frames and used in conjunction with spectral regression [9]. The Gabor and SIFT features were each reduced to 30 dimensions per video frame using this technique. These choices for sample and feature size were informed by previous work and motivated by the computational limitations imposed by each method.

2.3.4 Cross-validation

To prevent model over-fitting, stratified k-fold cross-validation [31] was used. Cross-validation procedures typically involve partitioning the data and iterating through the partitions such that all the data is used but no classifier iteration is trained and tested on the same data. Stratified cross-validation procedures ensure that the resultant partitions have roughly equal distributions of the target class (in this case AU12). This property is desirable because many performance metrics are highly sensitive to class skew [46] (e.g., F_1 cannot be calculated if the target class never occurs). By using the same partitions across methods, the randomness introduced by repeated repartitioning can also be avoided. In this study, each video segment was assigned to one of five partitions (called "folds"). For each iteration of the cross-validation procedure, three folds were used for training, one fold was used for validation (i.e, parameter optimization), and one fold was used for testing.

2.3.5 Prediction

Following previous work on binary expression detection [25, 81], support vector machines (SVM) [79] were used for binary classification. SVM classifiers apply the "kernel trick," which uses dot product, to keep computational loads reasonable. The kernel function (in this case, a radial basis function) enables the SVM to fit a hyperplane of maximum margin into the transformed high dimensional feature space. SVMs were trained using two classes corresponding to the FACS occurrence codes described in subsection 2.2.1 (i.e., 0 or 1). Training sets were created by randomly sampling 10,000 frames with roughly equal representation for each class. The choice of sample size was motivated by the computational limitations imposed by model training during cross-validation. Classifier and kernel parameters (i.e., C and γ, respectively) were optimized using a "grid-search" procedure [42] on a separate validation set. The output values of the SVM models were fractions corresponding to the distance of each frame's high dimensional feature point from the class-separating hyperplane. These values were used for smile intensity estimation and also discretized using the standard SVM threshold of zero to provide predictions for binary smile detection (i.e., negative values were labeled absence of AU12 and positive values were labeled presence of AU12).

Following previous work on expression intensity estimation using multiclass classifiers [66, 60, 62], the SVM framework was extended for multiclass classification using the "one-against-one" technique [41]. In this technique, if k is the number of classes, then $k(k-1)/2$ subclassifiers are constructed and each one trains data from two classes; classification is then resolved using a subclassifier voting strategy. Multiclass SVMs were trained using six classes corresponding to the FACS intensity codes described in subsection 2.2.2. Training sets were created by randomly sampling 10,000 frames with roughly equal representation for each class. Classifier and kernel parameters (i.e., C and γ, respectively) were optimized using a "grid-search" procedure [42] on a separate validation set. The output values of the multiclass classifiers were integers corresponding to each frame's estimated smile intensity level. These values were used for smile intensity estimation and also discretized to provide predictions for binary smile detection (i.e., values of 0 were labeled absence of AU12 and values of 1 through 5 were labeled presence of AU12).

Following previous work on expression intensity estimation using regression [47, 51, 73, 18, 50], epsilon support vector regression (ε-SVR) [79] was used. As others have noted [73], ε-SVR is appropriate to expression intensity estimation because its ε-insensitive loss function is robust and generates a smooth mapping. ε-SVRs were trained using six classes corresponding to the FACS intensity codes described in Section 2.2.2. Training sets were created by randomly sampling 10,000 frames with roughly equal representation for each class. Classifier and kernel parameters (i.e., C and γ, respectively) were optimized using a "grid-search" procedure [42]; the epsilon parameter was left at the default value (ε=0.1). The output values of the regression models were fractions corresponding to each frame's estimated smile intensity level. This output was was used for smile intensity estimation and also discretized using a threshold of 0.5 to provide predictions for binary smile detection (i.e., values that would round to 0 were labeled absence of AU12 and values that would round to 1 or above were labeled presence of AU12).

2.4 PERFORMANCE EVALUATION

The majority of previous work on expression intensity estimation has utilized the Pearson product-moment correlation coefficient (PCC) to measure the correlation between intensity estimations and ground truth coding. PCC is invariant to linear transformations, which is useful when using estimations that differ in scale and location from the ground truth coding (e.g., classifier margins). However, this same property is problematic when the estimations *are* similar to the ground truth (e.g., multiclass classifier predictions), as it introduces an undesired handicap. For instance, a classifier that always estimates an expression to be two intensity levels stronger than it is will have the same PCC as a classifier that always estimates the expression's intensity level correctly. For this reason, we performed our analyses using another performance metric that grants more control over its relation to linear transformations: the intraclass correlation coefficient (ICC) [77]. The ICC formula provided in Equation 2.1 (using Between-Target and Within-Target Mean Squares) was used for the multiclass SVM and ε-SVR classifiers as their outputs were consistent with that of ground truth

coding. The ICC formula provided in Equation 2.2 (using Between-Target Mean Squares and Residual Sum of Squares) was used for the SVM margin classifier because it takes into account differences in the scale and location of the two measures.

The majority of previous work on binary expression detection has utilized receiver operating characteristic curves. When certain assumptions are met, the area under the curve (AUC) is equal to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [26]. The fact that AUC captures information about the entire distribution of decision points is a benefit of the measure, as it removes the subjectivity of threshold selection. However, in the case of automatic expression annotation, a threshold *must* be chosen in order to create predictions that can be compared with ground truth coding. In light of this issue (and recent critiques of the AUC measure [35, 58]), we perform our analyses using a threshold-specific performance metric: the F_1 score, which is the harmonic mean of precision and recall (Equation 2.3) [78].

$$ICC(1,1) = \frac{BMS - WMS}{BMS + (k-1)WMS} \tag{2.1}$$

$$ICC(3,1) = \frac{BMS - EMS}{BMS + (k-1)EMS} \tag{2.2}$$

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{2.3}$$

2.5 DATA ANALYSIS

Main effects and interaction effects among the different methods were analyzed using two univariate general linear models [43] (one for binary smile detection and one for smile intensity estimation). F_1 and ICC were entered as the sole dependent variable in each model, and database, extraction type, reduction type, and classification type were entered as "fixed factor" independent variables. The direction of significant differences were explored using marginal means for all variables except for classification type. In this case, post-hoc Tukey HSD tests [43] were used to explore differences between the three types of classification.

3.0 RESULTS

3.1 SMILE INTENSITY ESTIMATION

Across all methods and databases, the average intensity estimation performance was ICC=0.64. However, performance varied widely between databases and methods, from a low of ICC=0.23 to a high of ICC=0.92 (Figure 5).

The overall general linear model for smile intensity estimation was significant (Table 1). The included independent variables accounted for 76.5% of the variance in intensity estimation performance. Main effects of database, extraction method, and classification method were apparent. Intensity estimation performance was significantly higher for the BP4D database than for the Spectrum database, and intensity estimation performance using SIFT features was significantly higher than that using Gabor features. Intensity estimation performance using Multiclass SVM and ε-SVR classification was significantly higher than that using SVM margin classification. There was no significant difference in performance between Laplacian Eigenmap and PCA for reduction.

These main effects were qualified by four significant interaction effects. First, the difference between SIFT features and Gabor features was greater in the Spectrum database than in the BP4D database. Second, while Laplacian Eigenmap performed better in the Spectrum database, PCA performed better in the BP4D database. Third, while Multiclass SVM performed better in the Spectrum database, ε-SVR performed better in the BP4D database. Fourth, PCA reduction yielded higher intensity estimation performance when combined with SVM margin classification, but lower intensity estimation performance when combined with Multiclass SVM and ε-SVR classification.

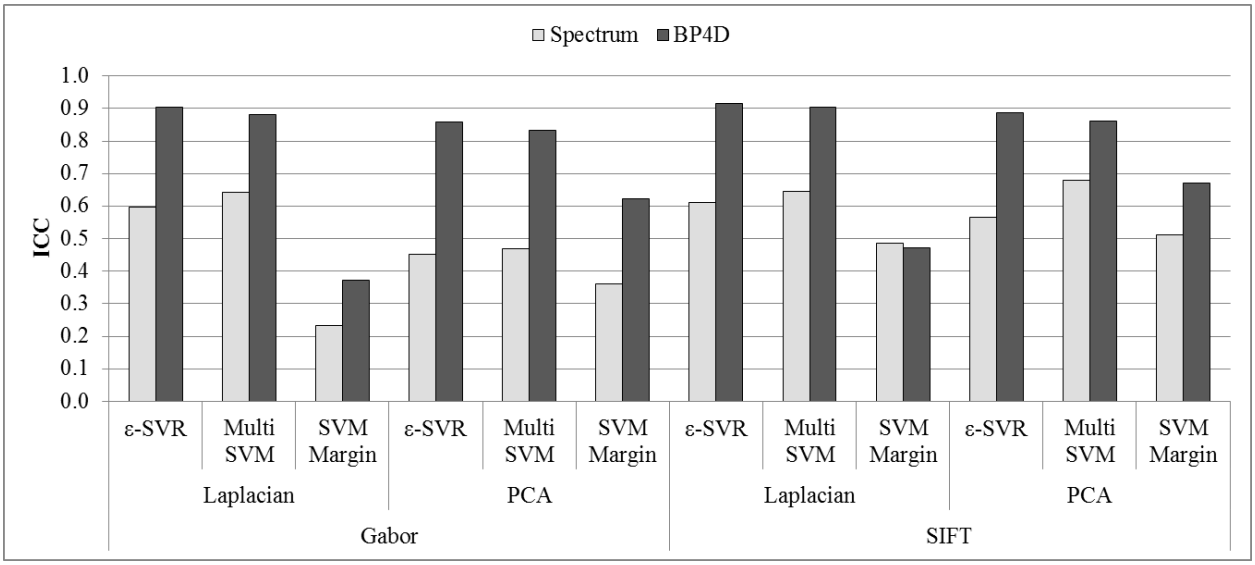

Figure 5: Smile Intensity Estimation Performance

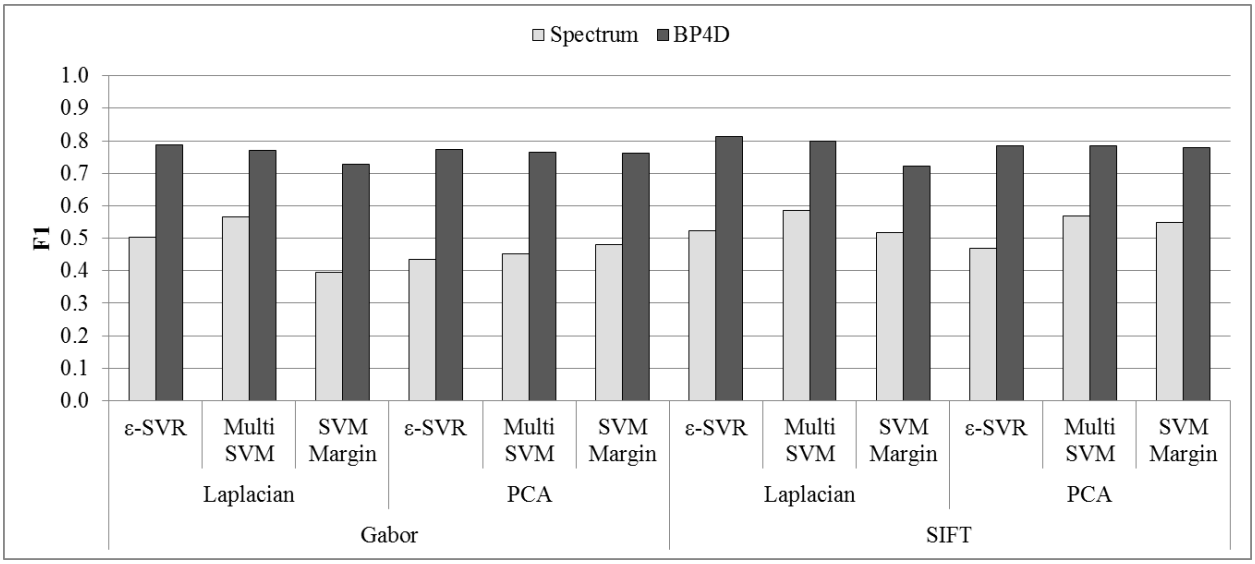

Figure 6: Binary Smile Detection Performance

Table 1: General Linear Model results for Smile Intensity Estimation (ICC)

	Spectrum	BP4D		F	p
Database	0.521	0.765		158.206	**.00**
	Gabor	SIFT		F	p
Extraction	0.602	0.684		17.892	**.00**
	Laplacian	PCA		F	p
Reduction	0.639	0.648		0.197	.66
	SVM Margin	Multi SVM	ε-SVR	F	p
Classification	0.467^{a}	0.739^{b}	0.724^{b}	83.360	**.00**
Interaction Effects				F	p
Database×Extraction				4.627	**.03**
Database×Reduction				3.958	**.05**
Database×Classification				8.873	**.00**
Reduction×Classification				13.391	**.00**

Note: Numbers with dissimilar superscripts are significantly different by Tukey HSD.

Table 2: General Linear Model Results for Binary Smile Detection (F_1)

	Spectrum	BP4D		F	p
Database	0.504	0.772		440.209	**.00**
	Gabor	SIFT		F	p
Extraction	0.618	0.658		9.740	**.00**
	Laplacian	PCA		F	p
Reduction	0.642	0.633		0.501	.48
	SVM Margin	Multi SVM	ε-SVR	F	p
Classification	0.616^{a}	0.661^{b}	0.636	4.175	**.02**
Interaction Effects				F	p
Reduction×Classification				5.753	**.00**

Note: Numbers with dissimilar superscripts are significantly different by Tukey HSD.

3.2 BINARY SMILE DETECTION

Across all methods and databases, the average binary detection performance was F_1=0.64. However, performance varied between databases and methods, from a low of F_1=0.50 to a high of F_1=0.81 (Figure 6).

The overall general linear model for binary smile detection was significant. The included independent variables accounted for 79.6% of the variance in detection performance (Table 2). Main effects of database, extraction method, and classification method were apparent. Detection performance on the BP4D database was significantly higher than that on the Spectrum database, and detection performance using SIFT features was significantly higher than that using Gabor features. Detection performance using Multiclass SVM was significantly higher than that using SVM margin classification.

These main effects were qualified by a significant interaction effect between reduction method and classification method. PCA reduction yielded higher detection performance when combined with SVM margin classification, but lower detection performance when combined with Multiclass SVM or ε-SVR classification.

4.0 DISCUSSION

4.1 SMILE INTENSITY ESTIMATION

Intensity estimation performance varied between databases, feature extraction methods, and classification methods. Performance was higher in the BP4D database than in the Spectrum database. It is not surprising that performance differed between the two databases, given how much they differed in terms of participant demographics, social context, and image quality. Further experimentation will be required to pinpoint exactly what differences between the databases contributed to this drop in performance, but we suspect that illumination conditions, frontality of camera placement, and participant head pose were involved. It is also possible that the participants in the Spectrum database were more difficult to analyze due to their depressive symptoms. Previous research has found that nonverbal behavior (and especially smiling) changes with depression symptomatology [33]. There were also differences between databases in terms of social context that likely influenced smiling behavior; Spectrum was recorded during a clinical interview about depression symptoms, while BP4D was recorded during tasks designed to elicit specific and varied emotions. Participants in the Spectrum database smiled less frequently (20.5% of frames) and less intensely (average intensity 1.5) than did participants in the BP4D database (56.4% of frames and average intensity 2.4). This difference may have affected the difficulty of smile intensity estimation.

More surprising was that intensity estimation performance was higher for SIFT features than for Gabor features. This finding is encouraging from a computational load perspective, considering the toolbox implementation of SIFT used in this study [80] was many times faster than our custom implementation of Gabor. However, it is possible that SIFT was particularly well-suited to our form of registration with dense facial landmarking. Although

we were unable to test this hypothesis in the current study, it would have been interesting to compare these two methods of feature extraction in conjunction with a method of registration using sparse landmarking (e.g., holistic face detection or eye tracking). It is also important to note that the difference between SIFT and Gabor features was larger in the Spectrum database than in BP4D, which may be evidence of SIFT's appellative invariance.

For dimensionality reduction, intensity estimation performance was not significantly different between Laplacian Eigenmap and PCA. This may be an indication that the features used in this study were linearly separable and that manifold learning was unnecessary. This finding is also encouraging from a computational load perspective, as PCA is a much faster and simpler technique. However, it is important to note that the success of each dimensionality reduction technique depended on the database and on the classification method used. Laplacian Eigenmap was better suited to the Spectrum database, multiclass SVM classification, and ε-SVR classification, while PCA was better suited to the BP4D database and SVM margin classification. This finding suggests that supervised and nonlinear dimensionality reduction techniques may be particularly useful for challenging databases.

Most relevant to our main hypothesis are the findings regarding classification method. In line with the notion that intensity estimation and binary detection are different problems requiring different solutions, the intensity-trained multiclass SVM and ε-SVR classifiers performed significantly better at intensity estimation than the margins of binary-trained SVM classifiers. It is perhaps not surprising that classifiers trained for a specific task outperformed a classifier that was re-purposed without changes or adaptations. However, contrary to this notion, the intensity estimation performance yielded by SVM margins was not negligible. In some databases and with some methods, performance was even admirable.

4.2 BINARY SMILE DETECTION

Binary detection performance also varied between databases, feature extraction methods, and classification methods. These differences were very similar to those for expression intensity estimation. Binary detection performance was higher for the BP4D database than for

the Spectrum database, higher for SIFT features than for Gabor features, and no different between Laplacian Eigenmap and PCA for dimensionality reduction.

Surprisingly, detection performance was higher for Multiclass SVM classification than for SVM margin classification. This suggests that the best classifier for binary detection is not necessarily the one trained on binary labels. Perhaps having subclassifiers trained on each intensity level resolved the ambiguous cases of low intensity expressions, which might fall closer to the distribution of an overall negative class than to that of an overall positive class. As far as we know, this is the first study to attempt binary expression detection using an intensity-trained classifier. Although collecting frame-level intensity ground truth is labor-intensive, our findings indicate that this investment is worthwhile for even binary expression detection. That ε-SVR classification was not significantly better than SVM margins may be a result of the discrepancy between levels of measurement for the output values and ground truth coding. Because the fractional output of the ε-SVR had to be discretized using an arbitrary threshold (i.e., 0.5), borderline cases may have hurt its performance.

4.3 CONCLUSIONS

The primary goal of this paper was to learn whether binary expression detection and expression intensity estimation are different problems that require different solutions or similar problems that can be approached with the same methods and even the same classifiers. Our results indicate that these two problems require similar methods and can even be approached using the same classifiers; however, these are not the classifiers that Bartlett et al. [4] had in mind. Rather than the margins of binary-trained classifiers, our results support the use of intensity-trained multiclass classifiers for both binary expression detection and expression intensity estimation. SIFT features are recommended over Gabor features in conjunction with dense facial landmarking, and unsupervised PCA is shown to be a competitive option for dimensionality reduction in the context of facial expression analysis. The fact that these results were replicated in two separate (and quite different) databases increases our confidence of their generalizability.

4.4 LIMITATIONS AND FUTURE DIRECTIONS

The primary limitation of the current study was that it focused on a single facial expression and did not test if these findings would generalize to others. Future work should definitely explore this issue by comparing different methods for the intensity estimation of other expressions. One limitation of the current study that might have influenced its results is a divergence between the number of reduced features yielded by Laplacian Eigenmap and PCA. Although we followed the standard practice for each dimensionality reduction technique, this difference may have contributed to our mixed findings on the topic. Finally, future work would benefit from a comparison of different registration techniques and cross-validation procedures. The examination of additional (and more varied) methods for feature extraction, dimensionality reduction, and classification would also be informative.

5.0 ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant CNS 1205195. Technical and programming support was provided by Mohammad H. Mahoor, S. Mohammad Mavadati, and Laszlo A. Jeni. FACS coding and management was provided by Nicki Siverling, Dean P. Rosenwald, and Shawn Zuratovic.

BIBLIOGRAPHY

[1] E. L. Abel and M. L. Kruger. Smile Intensity in Photographs Predicts Longevity. *Psychological Science*, 21(4):542–544, 2010.

[2] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (4th ed.)*. Washington, DC, 1994.

[3] L. F. Barrett, B. Mesquita, and M. Gendron. Context in emotion perception. *Current Directions in Psychological Science*, 20(5):286–290, 2011.

[4] M. S. Bartlett, G. Littlewort, B. Braathen, T. J. Sejnowski, and J. R. Movellan. A prototype for automatic recognition of spontaneous facial actions. In S. Becker and K. Obermayer, editors, *Advances in Neural Information Processing Systems*. MIT Press, 2003.

[5] M. S. Bartlett, G. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan. Automatic Recognition of Facial Actions in Spontaneous Expressions. *Journal of Multimedia*, 1(6):22–35, 2006.

[6] M. S. Bartlett, G. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan. Fully Automatic Facial Action Recognition in Spontaneous Behavior. In *IEEE International Conference on Automatic Face & Gesture Recognition*, pages 223–230, Southampton, 2006.

[7] J. B. Bavelas and C. N. Faces in dialogue. In J. A. Russell and J. M. Fernandez-Dols, editors, *The psychology of facial expression*, pages 334–346. Cambridge University Press, New York, 1997.

[8] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[9] D. Cai, X. He, W. V. Zhang, and J. Han. Regularized locality preserving indexing via spectral regression. In *Conference on Information and Knowledge Management*, pages 741–750. ACM, 2007. ISBN 1595938036.

[10] S. W. Chew, P. Lucey, S. Lucey, J. Saragih, J. F. Cohn, I. Matthews, and S. Sridharan. In the Pursuit of Effective Affective Computing: The Relationship Between Features and

Registration. *IEEE Transactions on Systems, Man, and Cybernetics*, 42(4):1006–1016, 2012.

[11] J. F. Cohn and K. L. Schmidt. The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(2):57–72, 2004.

[12] J. F. Cohn, T. S. Kruez, I. Matthews, Y. Ying, N. Minh Hoai, M. T. Padilla, Z. Feng, and F. De La Torre. Detecting depression from facial actions and vocal prosody. In *Affective Computing and Intelligent Interaction*, pages 1–7, Amsterdam, 2009.

[13] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[14] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80, 2001.

[15] C. Darwin. *The expression of emotions in man and animals*. Oxford University, New York, 3rd edition, 1872.

[16] J. G. Daugman. Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(7):1169–1179, 1988.

[17] O. Deniz, M. Castrillon, J. Lorenzo, L. Anton, and G. Bueno. Smile Detection for User Interfaces. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, F. Porikli, J. Peters, J. Klosowski, L. Arns, Y. K. Chun, T.-M. Rhyne, and L. Monroe, editors, *Advances in Visual Computing*, volume 5359 of *Lecture Notes in Computer Science*, pages 602–611. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[18] A. Dhall and R. Goecke. Group expression intensity estimation in videos via Gaussian Processes. *International Conference on Pattern Recognition*, pages 3525–3528, 2012.

[19] P. Ekman. Facial expressions of emotion: New findings, new questions. *Psychological Science*, 3(1):34–38, 1992.

[20] P. Ekman. Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, 1000(1):205–221, 2003.

[21] P. Ekman and E. L. Rosenberg. *What the Face Reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, 2nd edition, 2005.

[22] P. Ekman, W. V. Friesen, and S. Ancoli. Facial signs of emotional experience. *Journal of Personality and Social Psychology*, 39(6):1125–1134, 1980.

[23] P. Ekman, R. J. Davidson, and W. V. Friesen. The Duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology*, 58(2): 342–353, 1990.

[24] P. Ekman, W. V. Friesen, and J. Hager. *Facial Action Coding System (FACS): A technique for the measurement of facial movement.* Research Nexus, Salt Lake City, UT, 2002.

[25] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36(1):259–275, 2003.

[26] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8): 861–874, 2006.

[27] W. Fellenz, J. G. Taylor, N. Tsapatsoulis, and S. Kollias. Comparing template-based, feature-based and supervised classification of facial expressions from static images. *Computational Intelligence and Applications*, 1999.

[28] M. G. Frank, P. Ekman, and W. V. Friesen. Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology*, 64(1):83–93, 1993.

[29] A. J. Fridlund. Sociality of solitary smiling : potentiation by an implicit audience. *Journal of Personality and Social Psychology*, 60(2):12, 1991.

[30] A. J. Fridlund. The behavioral ecology and sociality of human faces. In M. S. Clark, editor, *Review of Personality Social Psychology*, pages 90–121. Sage Publications, 1992.

[31] S. Geisser. *Predictive Inference.* Chapman and Hall, New York, NY, 1993. ISBN 0-412-03471-9.

[32] J. M. Girard. CCS Coding Software, 2013. URL http://pitt.edu/~jmg174/ccs.

[33] J. M. Girard, J. F. Cohn, M. H. Mahoor, S. M. Mavadati, and D. P. Rosenwald. Social Risk and Depression: Evidence from manual and automatic facial expression analysis. In *IEEE International Conference on Automatic Face & Gesture Recognition*, 2013.

[34] M. Hamilton. Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, 6(4):278–296, 1967.

[35] D. J. Hand. Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, 77(1):103–123, June 2009. ISSN 0885-6125.

[36] L. Harker and D. Keltner. Expressions of Positive Emotion in Women's College Yearbook Pictures and Their Relationship to Personality and Life Outcomes Across Adulthood. *Journal of Personality and Social Psychology*, 80(1):112–124, 2001.

[37] M. Hertenstein, C. Hansel, A. Butts, and S. Hile. Smile intensity in photographs predicts divorce later in life. *Motivation and Emotion*, 33(2):99–105, 2009.

[38] U. Hess, R. Banse, and A. Kappas. The intensity of facial expression is determined by underlying affective state and social situation. *Journal of Personality and Social Psychology*, 69(2):280–288, 1995. doi: 10.1037/0022-3514.69.2.280.

[39] U. Hess, S. Blairy, and R. E. Kleck. The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *Journal of Nonverbal Behavior*, 24 (4):265–283, 2000.

[40] U. Hess, R. B. Adams Jr, and R. E. Kleck. Who may frown and who should smile? Dominance, affiliation, and the display of happiness and anger. *Cognition and Emotion*, 19(4):515–536, 2005.

[41] C.-W. Hsu and C.-J. Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(4):415–425, Jan. 2002. ISSN 1045-9227.

[42] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. Technical report, 2003.

[43] IBM Corp. IBM SPSS Statistics for Windows, Version 21.0, 2012.

[44] Image Metrics. LiveDriver SDK, 2013. URL http://www.image-metrics.com/.

[45] C. E. Izard. *The face of emotion*. Appleton-Century-Crofts, New York, 1971.

[46] L. A. Jeni, J. F. Cohn, and F. De La Torre. Facing imbalanced data: Recommendations for the use of performance metrics. In *International Conference on Affective Computing and Intelligent Interaction*, 2013.

[47] L. A. Jeni, J. M. Girard, J. F. Cohn, and F. D. L. Torre. Continuous AU Intensity Estimation using Localized, Sparse Facial Feature Space. In *International Workshop on Emotion Representation, Analysis and Synthesis in Continuous Time and Space*, 2013.

[48] J. P. Jones and L. a. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of neurophysiology*, 58(6):1233–1258, Dec. 1987. ISSN 0022-3077.

[49] G. Juckel, R. Mergl, A. Prassl, P. Mavrogiorgou, H. Witthaus, H. J. Moller, and U. Hegerl. Kinematic analysis of facial behaviour in patients with schizophrenia under emotional stimulation by films with "Mr. Bean". *European Archives of Psychiatry and Clinical Neuroscience*, 258(3):186–191, 2008. URL http://www.ncbi.nlm.nih.gov/pubmed/18071625.

[50] L. Ka Keung and X. Yangsheng. Real-time estimation of facial expression intensity. In *IEEE International Conference on Robotics and Automation*, volume 2, pages 2567–2572, 2003. ISBN 1050-4729.

[51] S. Kaltwang, O. Rudovic, and M. Pantic. Continuous Pain Intensity Estimation from Facial Expressions. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, C. Fowlkes, S. Wang, M.-H. Choi, S. Mantler, J. Schulze, D. Acevedo, K. Mueller, and M. Papka, editors, *Advances in Visual Computing*, volume 7432 of *Lecture Notes in Computer Science*, pages 368–377. Springer, Berlin, Heidelberg, 2012.

[52] D. Keltner. Evidence for the Distinctness of Embarrassment, Shame, and Guilt: A Study of Recalled Antecedents and Facial Expressions of Emotion. *Cognition and Emotion*, 10 (2):155–172, 1996.

[53] S. Koelstra and M. Pantic. Non-rigid registration using free-form deformations for recognition of facial actions and their temporal dynamics. In *IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–8, 2008. ISBN 1424421535.

[54] E. Krumhuber, A. S. Manstead, and A. Kappas. Temporal Aspects of Facial Displays in Person and Expression Perception: The Effects of Smile Dynamics, Head-tilt, and Gender. *Journal of Nonverbal Behavior*, 31(1):39–56, 2007.

[55] G. Littlewort, M. S. Bartlett, I. R. Fasel, J. Susskind, and J. R. Movellan. Dynamics of facial expression extracted automatically from video. *Image and Vision Computing*, 24 (6):615–625, 2006.

[56] G. Littlewort, M. S. Bartlett, and K. Lee. Automatic coding of facial expressions displayed during posed and genuine pain. *Image and Vision Computing*, 27(12):1797–1803, 2009.

[57] C. Liu, S. Louis, and H. Wechsler. A Gabor Feature Classifier for Face Recognition. In *IEEE International Conference on Computer Vision*, pages 270 – 275, 2001. ISBN 0769511430.

[58] J. M. Lobo, A. Jiménez-Valverde, and R. Real. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17(2): 145–151, Mar. 2008. ISSN 1466-822X.

[59] D. G. Lowe. Object recognition from local scale-invariant features. In *IEEE International Conference on Computer Vision*, pages 1150–1157 vol.2, 1999. ISBN 0-7695-0164-8.

[60] M. H. Mahoor, S. Cadavid, D. S. Messinger, and J. F. Cohn. A framework for automated measurement of the intensity of non-posed Facial Action Units. In *Computer Vision and Pattern Recognition Workshops*, pages 74–80, Miami, 2009.

[61] I. Matthews and S. Baker. Active Appearance Models Revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[62] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn. DISFA : A Spontaneous Facial Action Intensity Database. *IEEE Transactions on Affective Computing*, 2013.

[63] R. Mergl, M. Vogel, P. Mavrogiorgou, C. Gobel, M. Zaudig, U. Hegerl, and G. Juckel. Kinematical analysis of emotionally induced facial expressions in patients with obsessive-compulsive disorder. *Psychological Medicine*, 33(8):1453–1462, 2003.

[64] R. Mergl, P. Mavrogiorgou, U. Hegerl, and G. Juckel. Kinematical analysis of emotionally induced facial expressions: a novel tool to investigate hypomimia in patients suffering from depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 76(1): 138–140, 2005.

[65] D. S. Messinger, T. D. Cassel, S. I. Acosta, Z. Ambadar, and J. F. Cohn. Infant Smiling Dynamics and Perceived Positive Emotion. *Journal of Nonverbal Behavior*, 32(3):133–155, 2008.

[66] D. S. Messinger, M. H. Mahoor, S.-M. Chow, and J. F. Cohn. Automated Measurement of Facial Expression in Infant-Mother Interaction: A Pilot Study. *Infancy*, 14(3):285–305, 2009.

[67] C. Oveis, J. Gruber, D. Keltner, J. L. Stamper, and W. T. Boyce. Smile intensity and warm touch as thin slices of child and family affective style. *Emotion*, 9(4):544–548, 2009.

[68] I. S. Pandzic and R. Forchheimer. *The Origins of the MPEG-4 Facial Animation Standard*. Wiley Online Library, 2002.

[69] M. Pantic and L. J. M. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, 2003.

[70] K. M. Prkachin and P. E. Solomon. The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. *Pain*, 139(2):267–274, 2008.

[71] L. I. Reed, M. A. Sayette, and J. F. Cohn. Impact of depression on response to comedy: a dynamic facial coding analysis. *Journal of Abnormal Psychology*, 116(4):804–809, 2007.

[72] J. Reilly, J. Ghent, and J. McDonald. Investigating the dynamics of facial expression. *Advances in Visual Computing*, pages 334–343, 2006.

[73] A. Savran, B. Sankur, and M. Taha Bilge. Regression-based intensity estimation of facial action units. *Image and Vision Computing*, 2011.

[74] J. P. Seder and S. Oishi. Intensity of Smiling in Facebook Photos Predicts Future Life Satisfaction. *Social Psychological and Personality Science*, 3(4):407–413, 2012.

[75] K. Shimada, T. Matsukawa, Y. Noguchi, and T. Kurita. Appearance-Based Smile Intensity Estimation by Cascaded Support Vector Machines. In R. Koch and F. Huang, editors, *Computer Vision ACCV 2010 Workshops*, volume 6468, pages 277–286. Springer Berlin / Heidelberg, 2011.

[76] K. Shimada, Y. Noguchi, and T. Kurita. Fast and Robust Smile Intensity Estimation by Cascaded Support Vector Machines. *Internation Journal of Computer Theory and Engineering*, 5(1):24–30, 2013.

[77] P. E. Shrout and J. L. Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2):420, 1979.

[78] C. J. van Rijsbergen. *Information Retrieval*. Butterworth, London, 2nd edition, 1979.

[79] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.

[80] A. Vedali and B. Fulkerson. VLFeat: An Open and Portable Library of Computer Vision Algorithms, 2008. URL http://www.vlfeat.org/.

[81] J. Whitehill, G. Littlewort, I. R. Fasel, M. S. Bartlett, and J. R. Movellan. Toward Practical Smile Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):2106–2111, 2009.

[82] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1):37–52, 1987.

[83] P. Yang, L. Qingshan, and D. N. Metaxas. RankBoost with l1 regularization for facial expression recognition and intensity estimation. In *IEEE International Conference on Computer Vision*, pages 1018–1025, 2009. ISBN 1550-5499.

[84] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. 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.

[85] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, and P. Liu. A High-Resolution Spontaneous 3D Dynamic Facial Expression Database. In *IEEE International Conference on Automatic Face & Gesture Recognition*, Shanghai, 2013.