

2012

**University of North Carolina Wilmington
Master of Science in
Computer Science and Information Systems
Proceedings**

<https://csbapp.uncw.edu/mscsis>

DIFFERENTIATING DUCHENNE FROM NON-DUCHENNE SMILES USING ACTIVE
APPEARANCE MODELS AND THE FACIAL ACTION CODING SYSTEM

Jason Michael Vandeventer

A Thesis Submitted to the
University of North Carolina Wilmington in Partial Fulfillment
of the Requirements for the Degree of
Master of Science

Department of Computer Science
Department of Information Systems & Operations Management

University of North Carolina Wilmington

2012

Approved by

Advisory Committee

Curry Guinn

Bryan Reinicke

Eric Patterson

Chair

Accepted by

Dean, Graduate School

TABLE OF CONTENTS

ABSTRACT	iv
ACKNOWLEDGEMENTS	vi
DEDICATION	ix
LIST OF TABLES	x
LIST OF FIGURES	xi
1 INTRODUCTION	1
1.1 Overview	1
1.2 Problem Statement	5
1.3 Hypotheses	7
1.3.1 Hypothesis I	7
1.3.2 Hypothesis II	7
1.3.3 Hypothesis III	7
2 BACKGROUND MATERIAL	9
2.1 Facial Action Coding System	9
2.2 Smiles	11
2.3 Principal Component Analysis	15
2.4 Active Appearance Models	16
3 CLASSIFIERS	20
3.1 Overview	20
3.2 k -Nearest Neighbor	20
3.3 Naïve Bayes	21
3.4 Support Vector Machines	22
4 FACS AU-CODED IMAGE DATABASES	27
4.1 Cohn-Kanade	27

4.2	Bosphorus	30
5	PREVIOUS WORKS	33
5.1	Automatic Extraction of Action Units	33
5.2	Using AAMs for Expression Recognition	37
5.3	Posed Versus Spontaneous Smiles	42
6	EXPERIMENT	45
6.1	Overview	45
6.2	Data Collection	46
6.3	Methodology	48
7	RESULTS AND DISCUSSION	52
7.1	Results	52
7.2	Discussion	53
8	FUTURE WORK	58
	REFERENCES	62
	APPENDIX	77
A	Confusion Matrices	77

ABSTRACT

Affect recognition has been a popular area of study in computer vision and psychology over the past decade. According to the psychologist Dr. Paul Ekman, there are 7 “universal” emotions: happiness, sadness, surprise, fear, anger, disgust, and contempt. Happiness, for a long time, was considered to be any facial expression that contained a smile. Dr. Ekman showed that there are actually a number of different types of smiles, many of which are not linked to positive emotion, such as miserable smiles.

Duchenne smiles, which have been linked to genuine enjoyment and are therefore commonly referred to as *genuine* smiles, include the contraction of the orbicularis oculi, pars lateralis muscle, also known as the *Duchenne marker*. Non-Duchenne smiles lack this muscle contraction and are therefore commonly referred to as *posed* smiles. Many studies have attempted to differentiate Duchenne from non-Duchenne smiles using a variety of feature extraction and classification techniques. Many of these studies have used facial emotion expression images that have been labeled by certified Facial Action Coding System (FACS) coders.

In this study, the technique of using Active Appearance Models (AAMs) for feature extraction, and k -Nearest Neighbor (kNN), naïve Bayes, and Support Vector Machines (SVMs) for classification, to differentiate Duchenne

from non-Duchenne smiles is evaluated, using images from the Extended Cohn-Kanade AU-coded Facial Expression Database (CK+) and 2D images from the FACS AU-coded Bosphorus 3D Database (Bosphorus). This methodology has not been explored as of yet for the differentiation of Duchenne from non-Duchenne smiles, to the best of the author's knowledge. Three generalization approaches were evaluated: *ideal-case* (No generalization), *semi-generalized*, and *fully-generalized*. For all three approaches, kNN and naïve Bayes performed favorably to, and Support Vector Machines outperformed, previously published methodologies from similar studies. This indicates that Active Appearance Models may be a very strong choice for feature selection for affect recognition systems.

ACKNOWLEDGEMENTS

A special thank you to the members of my advisory board, all of whom have provided support and guidance.

A very special thanks to Dr. Curry Guinn for his support over the past 5 years. He encouraged me to join the Computer Science department, advised me on my first publication, and encouraged me to pursue graduate school. The quality of his support and guidance as an advisor is something I hope to emulate.

I'd like to thank Dali Hildebeitel, who, with her words and actions showed me what it means to be an excellent manager and a great person. Like the greatest sport coaches, she demanded the best from me every day and was always showing me how I could improve, in every facet of my life. She never settled for mediocrity and, for that, the respect I have for her is immeasurable. I learned so many valuable life lessons and grew both as an employee and as a person. For that, I am forever grateful.

To Dr. Caroline Clements and Dr. Richard Ogle. Thank you so very much for inviting me to work on the military research grant proposal. The amount of technical skills I developed while on the project is matched only by the amount of knowledge I gained about the proposal process. The balance struck between hard-work and enjoyment, in regards to your careers, is astounding.

The project was truly an amazing experience and to have helped secure a grant for you and the UNCW psychology department is something I will never forget.

A sincere thank you to Dr. Ann Stapleton and everyone from the iPlant Collaborative. The experience of working on a large-scale, NSF-funded project with some of the most brilliant researchers in the fields of computer science and biology gave me a new perspective on research and large-scale projects. To be associated with such an amazing project, with wonderful people, is truly an honor.

A special thank you to Dr. Ron Vetter: a professor I highly respect, for his passion for teaching, strong work-ethic, and concern for student success. He demands perfection and is always there to help a student achieve it. Dr. Vetter is also the professor who offered me the opportunity to work for the psychology, biology, foreign language departments, all of which were incredible experiences.

Thank you to Emma Kay Thornton for the innumerable things she has helped me with over the past 5 years. I could always count on her cheerful demeanor to add a smile to my face.

To Dr. Bryan Reinicke for being the epitome of the professor I hope to be. Thank you.

Thank you to Dr. Karl Ricanek, Jr.; Research Assistant Benjamin Bar-

bour, and everyone at the Institute for Interdisciplinary Studies in Identity Sciences (IISIS) for their expertise and support during my thesis process.

Finally, a thank you to everyone who has supported me along the way.

DEDICATION

To my parents, Debbie and Larry

Supporters; Critics

Experts; Novices

My Home; My Vacation

My Inspiration Emanates from their Dedication

LIST OF TABLES

1	Area under the ROC Curve Score (2-Alternative Forced Choice)	52
2	Comparison of similar published approaches. ¹ denotes Duchenne vs non-Duchenne. ² denotes posed vs. spontaneous study. (Note: If a study used multiple techniques, only the best performing combination and performance score is listed)	54
3	<i>Ideal-Case</i> Approach - Confusion Matrix - <i>k</i> -Nearest Neighbor	77
4	<i>Ideal-Case</i> Approach - Confusion Matrix - Naïve Bayes	77
5	<i>Ideal-Case</i> Approach - Confusion Matrix - Support Vector Machine	77
6	<i>Semi-Generalized</i> Approach - Confusion Matrix - <i>k</i> -Nearest Neighbor	78
7	<i>Semi-Generalized</i> Approach - Confusion Matrix - Naïve Bayes	78
8	<i>Semi-Generalized</i> Approach - Confusion Matrix - Support Vector Machine	78
9	<i>Fully-Generalized</i> Approach - Confusion Matrix - <i>k</i> -Nearest Neighbor	79
10	<i>Fully-Generalized</i> Approach - Confusion Matrix - Naïve Bayes	79
11	<i>Fully-Generalized</i> Approach - Confusion Matrix - Support Vector Machine	79

LIST OF FIGURES

1	FACS Intensity Scale	10
2	The zygomatic major is indicated by the <i>12</i> label in the left-hand picture. In the right-hand picture, the outer portion of the muscle, indicated by the <i>6</i> label, is the orbicularis oculi, pars lateralis. [34]	13
3	Neutral expression, Non-Duchenne Smile, and Duchenne Smile [34]	14
4	Example of a 227-point landmark scheme using an image from the CK+ database [54]	17
5	The PDM, the average point for each cluster, and the resulting average shape (from this study)	18
6	An landmarked image from the Bosphorus database, the corresponding triangulation, and the shape-free texture [65]	19
7	Example of a Linear SVM [10]	24
8	Example of a Gaussian RBF SVMs as gamma increases (C is held constant) [10]	25
9	Example of an image from the CK+ Database [54]	28
10	Example of an image from the Bosphorus Database [65]	31

11	The non-Duchenne, Duchenne, and combined (non-Duchenne and Duchenne) average face.	49
12	The original image, the reconstructed image from the <i>ideal-case</i> approach, and the reconstructed image from the <i>fully-generalized</i> approach. The quality of the reconstruction is high, which shows that the AAM, with both approaches, is accurate statistically in representing the original image.	51
13	Classifier performance given generalization approach	53

INTRODUCTION

Overview

Understanding human emotion is an invaluable skill. Emotions are the psychophysiological states that an individual can experience, such as happiness, sadness, anger, fear, disgust, contempt, and surprise. Understanding and being able to recognize emotion is important for interacting with others. Emotions are expressed in a variety of ways including body gestures, audible noises, and facial expressions.

Dr. Paul Ekman is a pioneer in the psychological study of human facial expressions. Dr. Ekman has traveled the world studying facial expressions from individuals in many diverse cultures. In the 1960's Dr. Ekman made the claim, against the popular belief at the time, that facial expressions are universal [31]. Many years later his research and the research of others in the field of psychology and anthropology continue to support this claim. Prior to his research, the popular belief was that facial expressions were cultural. This deduction resulted from the observation that individuals from different cultures had different reactions to similar stimuli. Dr. Ekman explained this finding by what he calls *display rules*, which are emotions that are so-

cially acceptable to display. These display rules are often culturally different and can dictate what, when, and to whom emotional expressions can be shown. A well-known example of display rules is an experiment which showed that Japanese and American students both displayed negative emotions when watching films about surgery and accidents. However, when a scientist sat down in the room with the students as they watched the films, the Japanese students masked their negative emotions with a smile while the American students did not. The basic emotion felt was the same, but the Japanese students had manipulated their expression to conform to their social norms. An involuntary display of the true emotion being experienced is known as *leakage*. The most common way leakage is observed is through *microexpressions*. Microexpressions are very fast facial movements that last around one-fifth of a second. Microexpressions occur during high-risk scenarios, such as when one is lying to an authority figure, and can give clues to the actual emotion experienced.

In 1978 Dr. Ekman and Dr. Wallace Friesen developed the Facial Action Coding System (FACS) [34]. FACS is a tool used to measure and describe facial movements and has become the de facto standard for measuring and describing facial expressions. The face is capable of making over 10,000 expressions. Not all expressions are linked to emotions. Using FACS, emotions and their intensities can be described. A FACS-trained individual can spot

expressions or microexpressions and use the information to make decisions about the individual displaying them. Many employers, such as law enforcement and government agencies, use FACS to train their employees in order to increase effectiveness by increasing their ability to detect deception. Humans are susceptible to error for a variety of reasons, such as stress and fatigue. Therefore, an interest in computer-based systems that can detect expressions exists.

Detecting human emotion as displayed by facial expressions is an important area in the field of Human-Computer Interaction (HCI) and Artificial Intelligence (AI). For computers to have the ability to intelligently interact with humans, they must be capable of recognizing and appropriately reacting to emotion. This not only includes interactions like those of HAL in *2001: A Space Odyssey* (although one could argue that HAL's ability to assess and "feel" human emotion is what led to the untimely demise of the crew); it may also include observing a driver's level of alertness or assessing a student's level of understanding during lectures. There are many useful applications for computer systems that can recognize human emotion, and one of the most pressing and interesting areas of research is deception detection.

Deception detection refers to the ability to recognize human expressions that are inconsistent with expected emotions. Deception occurs when an individual inhibits the expression of the emotion being experienced in order

to persuade observers that they are experiencing a different emotion. An individual may or may not be consciously aware that they are inhibiting an emotional expression. Through observations of leakage, an individual can spot these inconsistent expressions. While these observations give insight about deception, they do not guarantee the individual is being deceptive for a specific reason. For example, when an individual is asked by an authority figure if he stole an item, the individual may show guilt or fear. This does not prove that the guilt or fear is related to the crime in question. The individual may be feeling guilty about withholding the identity of the real thief, or he could fear that his truthful answer will not be believed. In many cases it is not sufficient to know *if* someone is being deceptive, but rather *why* they are being deceptive. Functional deception detection is therefore context-specific. The context in which the emotions are occurring is important and only through substantial observations can one make a decision about why someone is being deceptive.

Since deception detection is context-specific, a completely autonomous system may not be achievable. Regardless, the first step will be to develop a system that can aid in deception detection. For instance, a user may set the expected emotion (or emotions), and the computer system will flag any unrelated emotions it observes. In order to achieve this, a system must have the ability to accurately identify human facial emotions and microexpressions.

It is easy for most people to understand human emotion. Just by observing an individual's face, many people can, without specialized training or knowing the scientific reasons behind facial expressions, understand the emotion an individual is feeling (or pretending to feel). A computer does not have the ability to simply "look" at a human face. A simplified example of what a computer must do to "see" a human face is as follows: It must first locate the face in an image, determine the head position, extract features from the face, and classify those features. This process becomes more complex when a sequence of images is to be examined, or the face contains occlusions [43]. Automatically detecting a face or multiple faces in an image, extracting features and determining what they represent, and differentiating between posed and spontaneous facial actions are three of the main areas of research in the last ten years.

Problem Statement

Accurate recognition of subtle emotional differences in facial expressions is a major research area in computer science and psychology. Detecting facial emotions and differentiating them from other emotions, as well as determining the genuineness of the emotions, have been common areas of research over the past decade. The main challenges facing researchers have been defining

what constitutes an emotion, finding the best ways to extract features from facial images, how to classify those features, and determining if there is any discerning information in the features about subtle emotional differences in the subject.

The “Background Material” section discusses the Facial Action Coding System (FACS) which is a system that was designed to define facial actions. Facial emotions can be defined in terms of FACS *Action Units* (AUs). Different types of smiles and their subtle differences are also discussed. Active Appearance Models, a feature extraction technique, is discussed in detail before covering three machine learning classifiers. These classifiers are used in this study to determine if there is any discernible information regarding the differences between Duchenne and non-Duchenne smiles, based on the features extracted. Section 4 introduces two popular FACS AU-coded facial emotion databases which were used in this study. Section 5 contains a review of previous approaches to the problem of detecting and classifying facial emotions. Section 6 discusses the approach taken in this study, including the data collection process and experimental methodology. Section 7 covers the experiment results and a discussion of those results. Section 8 concludes this work by suggesting potential future research, based on the knowledge gained from this study.

Hypotheses

Hypothesis I

Using Active Appearance Models for feature extraction and k -Nearest Neighbor, Naïve Bayes, and Support Vector Machines for classification of Duchenne versus non-Duchenne smiles will classify these expressions correctly better than a “random-chance” classifier.

Hypothesis II

Using Active Appearance Models for feature extraction and k -Nearest Neighbor, Naïve Bayes, and Support Vector Machines for classification of Duchenne versus non-Duchenne smiles, and using all images for both the AAM model and classification phase, will result in performance better than other published methodologies from similar studies.

Hypothesis III

Using Active Appearance Models for feature extraction and k -Nearest Neighbor, Naïve Bayes and Support Vector Machines for classification of

Duchenne versus non-Duchenne smiles, a semi-generalized and fully generalized approach will result in performance that will compare favorably with other published methodologies from similar studies.

BACKGROUND MATERIAL

Facial Action Coding System

In the 1960's Dr. Paul Ekman made the claim, against the popular belief at the time, that facial expressions are universal [31]. Many years later his research and the research of others in the field of psychology and anthropology continues to support this claim. In order to objectively describe facial movement he developed the Facial Action Coding System, alongside Dr. Wallace Friesen, in 1978 [34]. The Facial Action Coding System, or FACS, is a system for precisely describing facial expressions using *Action Units* (AUs). Action Units roughly describe independent muscle movements. For instance, AU 12 represents the movements of the zygomatic major muscle which pulls the corners of the mouth upwards, obliquely. FACS has become the de-facto standard for measuring facial movements.

FACS consists of 44 action units, 14 head-movement codes, and 11 eye-movement codes. Action descriptors involve a group of muscles, such as the forward thrusting of the jaw, and are also described in FACS. AU's roughly represent the underlying muscles of the face and are distributed in such a way that divides the face into anatomically separate units and visually dis-

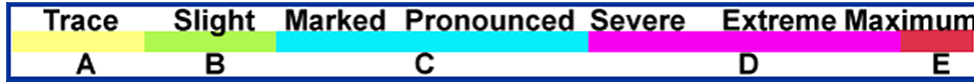


Figure 1: FACS Intensity Scale

tinguishable features. The face can produce more than 10,000 expressions. Any facial movement can be described in terms of action units or a combination of action units.

There are five aspects of facial movement that are measured: intensity, laterality, location, timing, and smoothness. The intensity of an action unit is described on a 5 point scale of A-E, which ranges from *trace* to *maximum*. Laterality, also known as the degree of symmetry, of the expression is scored on a 6 point scale. Location is the moment in time when each action unit begins and ends. Timing consists of the onset, apex, and ending time of an action unit, along with the span of time in between each temporal phase. Smoothness refers to whether or not the movement was smooth or irregular. All of these aspects make up the scoring of a facial movement [37]. The de-facto standard for accurate annotation of an image is the agreement of at least two FACS certified coders.

It takes approximately 100 hours of training to become a proficient FACS coder. FACS certification requires an individual to obtain at or greater than 80% agreement on a final test. The coding process is tedious. It can take around two hours for a FACS expert to code one minute of video [3, 40].

Littlewort et al. reported that human coders in their study could work 2 hours per day before negative effects were observed in accuracy and coder burn-out [51]. This is a major limitation of FACS and is why a major focus of the last ten years has been automating FACS coding.

FACS only identifies facial movements, not emotional expressions, in terms of action units. Ekman and Friesen developed the Emotion Facial Action Coding System (EMFACS) to address the need for emotion recognition using FACS. It describes facial expressions of emotion in terms of action unit combinations. There are currently 7 “universal” emotions defined using FACS action units. They are happiness, sadness, surprise, fear, anger, disgust, and contempt.

Smiles

There are over 18 different types of smiles, all with their own unique motivations, related to both positive and negative emotions [32]. Felt, false, miserable, and masking smiles are some of the most common examples. Felt smiles, also known as genuine smiles, are spontaneous expressions of positive emotion. False smiles, also known as deliberate smiles, attempt to convey positive emotion when none is actually felt. Miserable smiles convey a feeling of resigned acceptance of misery. Masking smiles, also known as concealment

smiles, attempt to conceal strong negative emotions by conveying positive emotions [37].

The muscles involved in each type of smile vary. Felt smiles consist of the zygomatic major pulling the lip corners upwards towards the cheekbone, and the outer portion of the orbicularis oculi (*orbicularis oculi, pars lateralis*) causing the skin around the eye socket to bunch. In FACS this is represented by AU 12 and AU 6, respectively. AU 12 is known as *Lip Corner Puller* and AU 6 is known as *Cheek Raiser*. This smile has been given the name *Duchenne smile* by Ekman, in honor of Duchenne de Boulogne, the man credited with first defining the muscle movements [33]. Duchenne smiles have been linked to genuine enjoyment and are also known as felt, enjoyment, and genuine smiles [38, 33, 35, 39, 70]. Therefore, smiles with only AU 12 are called *non-Duchenne* smiles. Non-Duchenne smiles are also known as unfelt, non-enjoyment, and false smiles. AU 6 has been termed *Duchenne Marker* because it distinguishes Duchenne from non-Duchenne smiles.

Before the disambiguation of smiles by Ekman and Friesen in the 1980's, many studies failed to observe the correlation of smiles with lying. These studies treated all smiles as a single expression of positive emotion. Many studies since then have been able to distinguish attributes of different smiles and define their relation to truthfulness [33].

In 1988 Ekman, Friesen, and O'Sullivan found that when enjoyment was

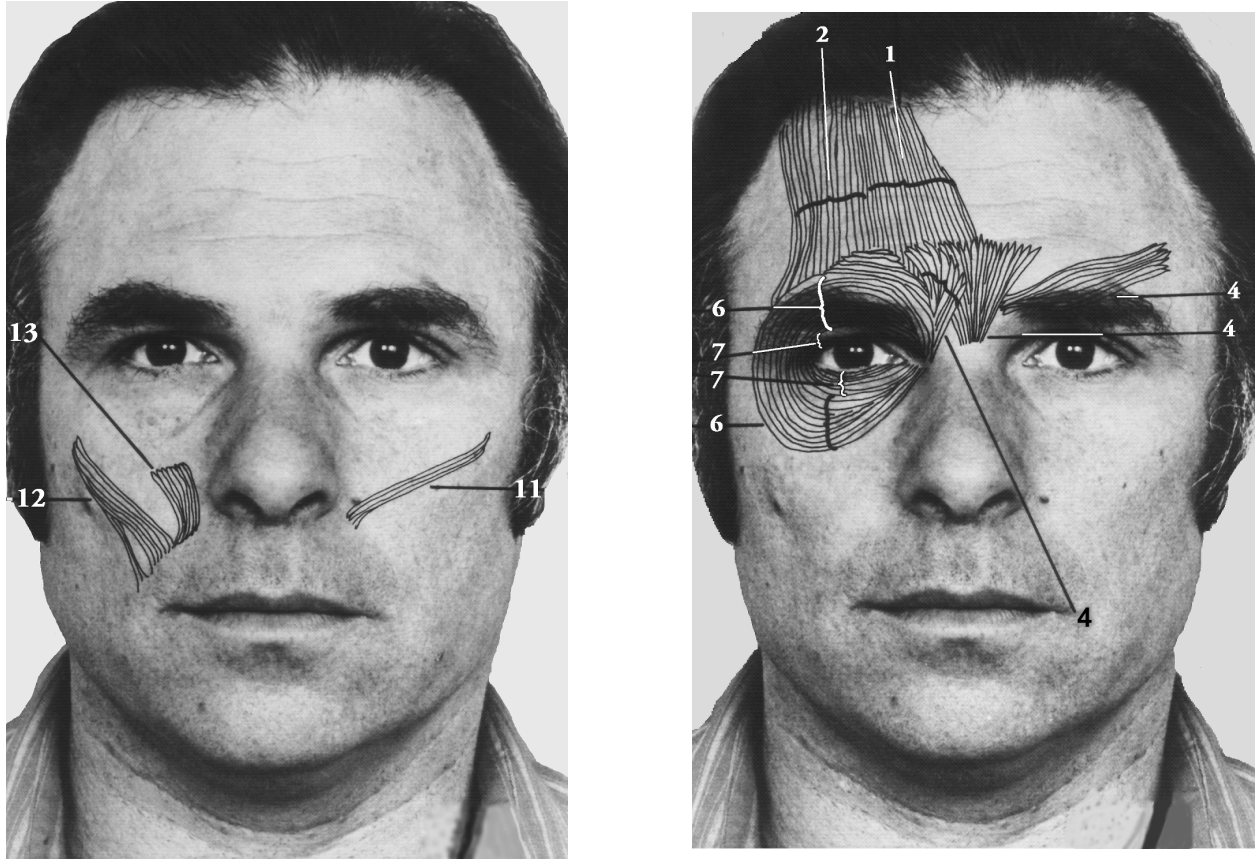


Figure 2: The zygomatic major is indicated by the 12 label in the left-hand picture. In the right-hand picture, the outer portion of the muscle, indicated by the 6 label, is the orbicularis oculi, pars lateralis. [34]

actually experienced by a subject, their smiles included activity of the orbicularis oculi, pars lateralis and this activity occurred more often than when enjoyment was feigned [38, 35]. This is consistent with prior findings that the presence of orbicularis oculi, pars lateralis activity in a smile is likely to indicate that it is felt, rather than deliberate. In a later experiment by Ekman, Davidson, and Friesen, similar results were achieved. They determined that Duchenne smiles, those smiles with the presence of both the zygomatic

major and orbicularis oculi, pars lateralis muscles (AU 6) were different from non-Duchenne smiles in that they are stronger indicators of positive emotions, such as enjoyment and happiness [33]. Cerebral activity from EEG recordings, facial EMG recordings, reactions of 10-month-old infants being approached by their mother rather than a stranger, and results from many other studies continue to support the claim that Duchenne smiles are strong indicators of genuine enjoyment and non-Duchenne smiles are strong indicators of posed enjoyment.[33, 39, 70]. It is important to note that the Duchenne marker does not guarantee genuine positive emotion. There is evidence that AU 6 can occur in other expressions, such as embarrassment smiles, and that some individuals can voluntarily activate the orbicularis oculi, pars lateralis muscle (AU 6) [1, 47].

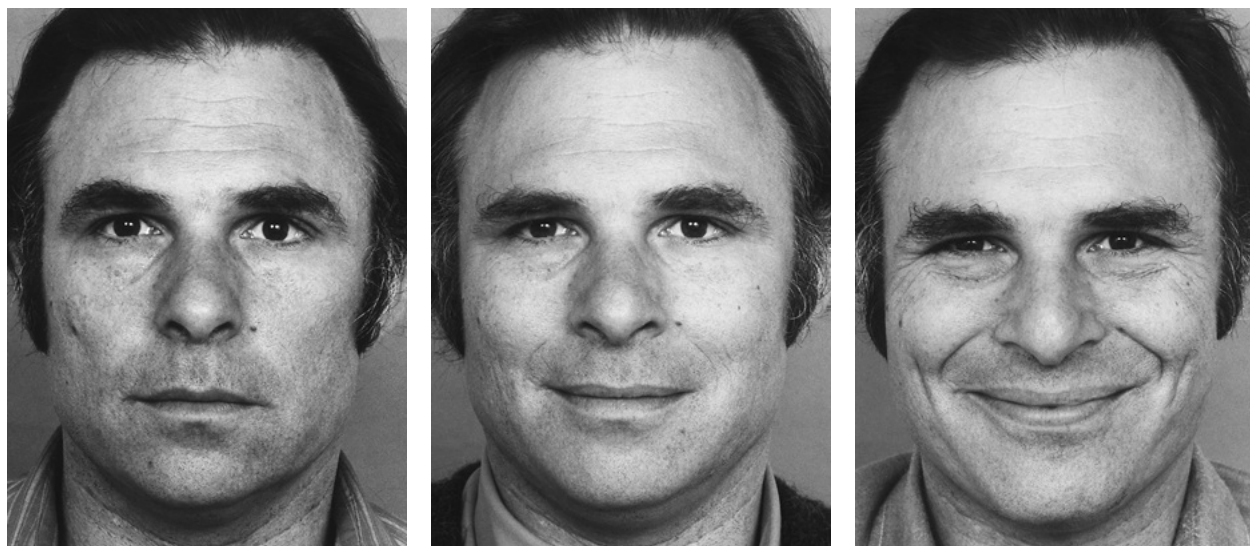


Figure 3: Neutral expression, Non-Duchenne Smile, and Duchenne Smile [34]

There are many social implications regarding Duchenne versus non-Duchenne smiles. In an experiment by Frank et al., individuals who displayed Duchenne smiles were perceived as being more positive [41]. Many factors contribute to smile genuineness and the perception of genuineness. Studies have shown that temporal dynamics in facial expressions are an important factor in both the perception of smile genuineness and the recognition of spontaneous facial actions. [1, 46, 72, 73, 76, 78].

Principal Component Analysis

Principal Component Analysis (PCA) was developed by Karl Pearson in 1901 and is a popular technique for reducing dimensionality in data using orthogonal transformation [59]. Data is converted from potentially correlated variables to a set of uncorrelated variables, called *principal components*, that represent the level of variance in the data. Principal components are ordered by decreasing level of variance [69]. This ordering allows the number of principal components to be truncated, while still maintaining the most significant information. Each principal component removed correlates to the removal of a single dimension. Reducing data dimensionality, and thus complexity, while retaining necessary information is one of the main reasons PCA is a popular technique.

Active Appearance Models

Active Appearance Models (AAMs) is a PCA-based statistical approach to modeling appearance developed by Cootes, Taylor, and Edwards in the late 1990's [22]. AAMs build on the idea of the Active Shape Models (ASMs), which is a model-based statistical approach to locating shapes in an image [24]. Active Appearance Models use annotated images where key areas are labeled using landmark points. Landmark schemes vary by the number and location of points. For a specific AAM the landmark scheme used should remain consistent across all images in the dataset. Active Appearance Models include both shape and texture information by combining a model of shape variation with a model of texture variation. Texture refers to the pattern of intensities or the colors in an image [25].

A Point Distribution Model (PDM) is developed using landmarked images from a training set [23]. The points in the landmarked image compose the shape of an object. Procrustes analysis is used to align the points of each image in the training set in a coordinate frame, with each image's points being represented by a vector, x [25]. For each point cluster, the average is computed and results in an average shape and, together, an average shape

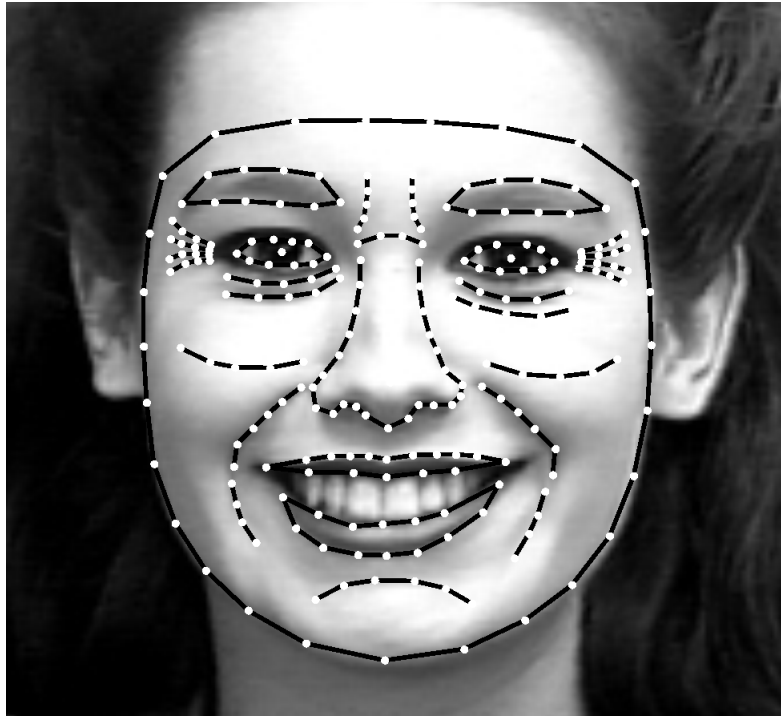


Figure 4: Example of a 227-point landmark scheme using an image from the CK+ database [54]

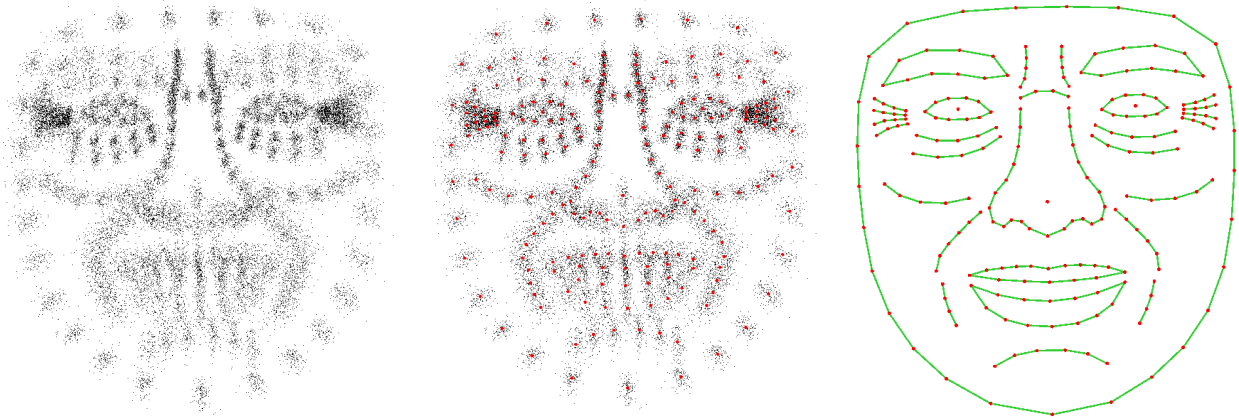


Figure 5: The PDM, the average point for each cluster, and the resulting average shape (from this study)

model of the PCA.

Once the PDM has been created, each image in the training set is warped to the average shape. Using triangulation on both a training image and the average shape model, where each triangle in the original training image corresponds to a triangle in the average shape model, the original image is warped to the average shape using piecewise affine warping [56]. This results in a *shape-free* texture image. Once this has been completed for all of the images in the training set, it is easy to calculate the average texture across the training set, resulting in a texture model. PCA is then performed on the shape and texture models. The combined appearance model is computed by concatenating the resulting shape and texture vectors, and performing PCA a final time to account for possible shape-texture correlations.

The combined appearance model can be used to interpret a previously

unseen image. A test image can be explained statistically by its relation to the appearance model in a vector of feature weights. Feature weights define a test image in terms of features existent in the appearance model. If the features of a test image are well-represented by the training set, a test image can be accurately expressed. That is, any feature that exists in the training set can be represented statistically. For faces, this means that even if an individual with a specific combination of features does not exist in the training set, a test image can be faithfully expressed statistically as long as those features exist across the training set.

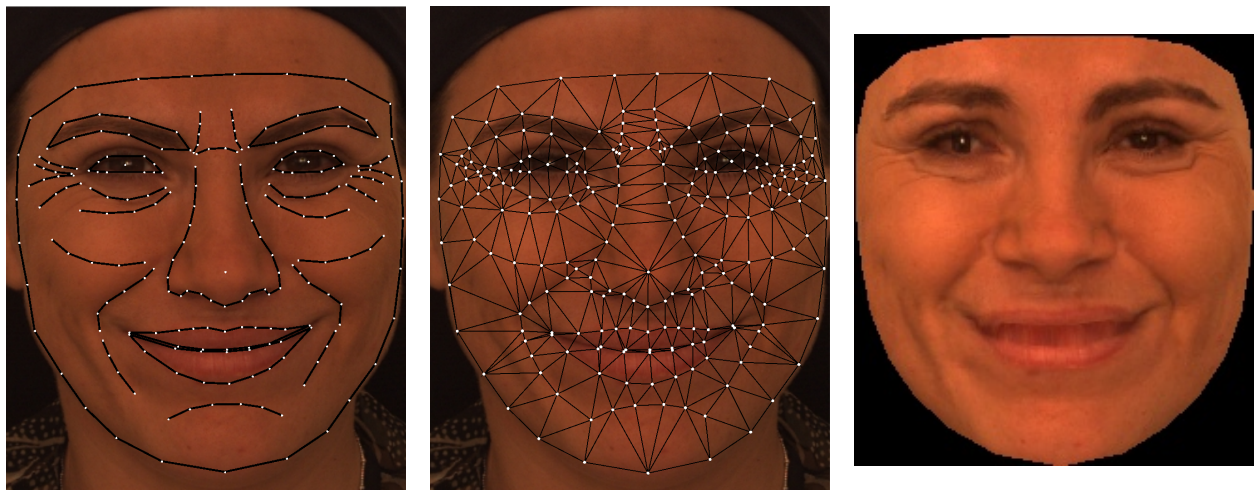


Figure 6: An landmarked image from the Bosphorus database, the corresponding triangulation, and the shape-free texture [65]

CLASSIFIERS

Overview

Two approaches exist for machine learning: supervised and unsupervised. Supervised learning algorithms use labeled training data which consist of vectors, each with a list of attributes and a target value, to make inferences about the data and to develop a function to accurately classify new data [29]. This function is tested on previously unseen data, typically known as the test set, to determine the classifier's accuracy. Popular supervised learning techniques include Feedforward neural networks and Support Vector Machines. Unsupervised learning algorithms do not use class-labeled data and are used to find similarities in data. Unsupervised learning algorithms are popular for use in clustering and dimensionality reduction problems using techniques such as k -means clustering and Principal Component Analysis.

k -Nearest Neighbor

k -Nearest Neighbor (k NN) is a supervised machine learning algorithm that classifies an object based on its proximity to training objects in a feature

space. When $k = 1$, the given object is assigned the class of the closest training object, or *neighbor*. When $k > 1$, the k -closest training objects classes are observed and the majority class is selected and assigned to the given object. Common distance metrics used for this algorithm include Euclidean, Mahalanobis, and Hamming [79].

Naïve Bayes

Naïve Bayes is an efficient supervised machine learning algorithm based on Bayes' theorem [9]. Feature attributes are assumed to be independent, given the class variable. While this assumption is rarely true in real-world applications, in practice, naïve Bayes performs well against more sophisticated classifiers [81]. Rish et al. showed that naïve Bayes performs best when features are either completely independent or functionally dependent [61].

Bayes' Theorem:

$$P(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)P(\omega_j)}{p(\mathbf{x})}$$

Naïve Bayes classifies an object using two steps: training and prediction. Training the classifier forms a probability model based on independent class features of a training set using supervised learning. The classifier assumes each class feature contributes independently to the class' probability. This class feature independence allows for quick training and greater generalization in class prediction. Once trained, the classifier can compute the subsequent probability of a test sample belonging to each class, classifying the sample based on the highest resulting probability.

Support Vector Machines

The Support Vector Machine (SVM) is a supervised machine learning algorithm used for data classification and regression analysis. SVMs produce a model based on data from a training set and then use that model to determine the class of previously unseen data, known as the test set. SVMs have traditionally been used for binary classification, but multi-class problems can be reduced into multiple binary classification problems using a multi-class SVM. For binary classification problems two approaches exist. Linear SVMs are used for linearly separable data, and nonlinear SVMs are used for non-linearly separable data. SVMs take training data and map it to a higher dimensional space. Linear SVMs attempt to find the hyperplane with the

maximal margin between classes in this space. The hyperplane that maximizes the margin between two classes is known as the optimal hyperplane [26]. A class boundary is defined by the point closest to the optimal hyperplane. These data points are known as *support vectors*. It is possible, for reasons such as mislabeled data, that the optimal hyperplane may not separate data without errors. For this reason, Cortes et al. introduced *soft margins* [26]. Soft margin hyperplanes allow the data to be separated with a minimal number of errors using a penalty function, C , also known as the *soft-margin constant*. By modifying C , a balance between the complexity of the decision rule and the frequency of error can be obtained.

SVMs were originally developed to solve linearly separable problems and were modified to solve nonlinear classification problems using a technique introduced by Boser et al. [12]. Kernels, such as polynomial and Radial Basis Function (RBF), can be applied to an SVM for nonlinear classification. In nonlinear classification, a transformed feature space is used to find the optimal hyperplane.

Four basic kernel functions exist for SVMs: Linear, polynomial, Radial

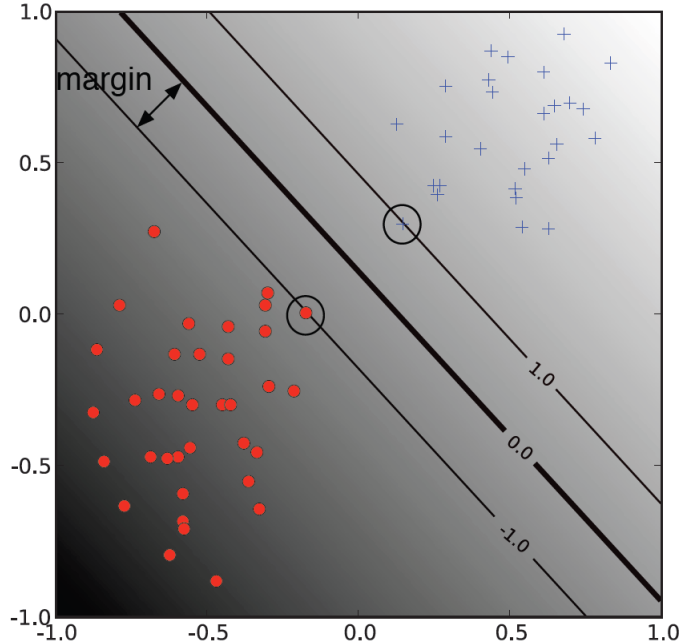


Figure 7: Example of a Linear SVM [10]

Basis Function (RBF), and sigmoid. These kernels take two parameters, C (soft margin constant, $C > 0$) and γ (gamma). Optimal values for C and γ vary depending on the given problem and few techniques exist for determining optimal values. One popular approach is a grid-search on C and γ using cross-validation. Combinations of C and γ are tried, and the combination with the best cross-validation accuracy is chosen. Using the C parameter for classification is common, but there is a newer approach that has become very popular. It replaces the C parameter with a ν parameter, and, thus, is aptly called the ν -soft margin support vector classifier, or ν -SVC (nu-SVC). $\nu \in [0, 1]$ and is related to the ratio of support vectors and the

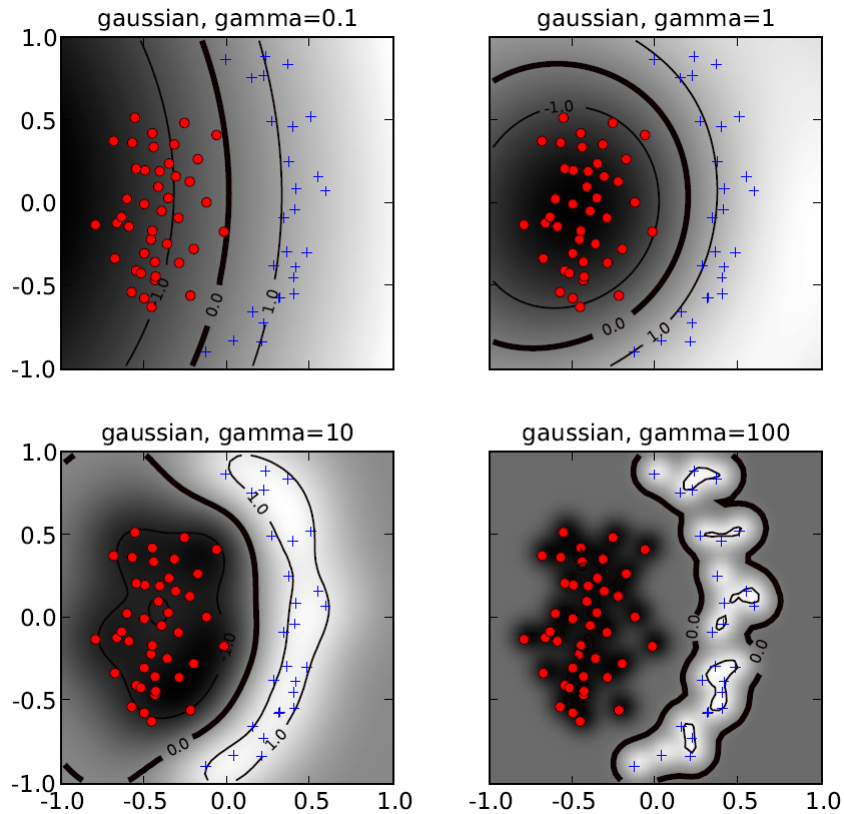


Figure 8: Example of a Gaussian RBF SVMs as gamma increases (C is held constant) [10]
ratio of the training error [67, 17, 18].

RBF kernels are widely used for a variety of reasons. They are preferred over linear kernels in cases where data is not linearly separable. They have the advantage over polynomial kernels because the polynomial kernels have a large number of hyperparameters, which can increase complexity. Polynomial kernels also have the ability to go to infinity or zero when the degree is large, whereas RBF kernels do not. RBF kernels are also often preferred over sigmoid kernels because under some parameters sigmoid kernels are not valid [77]. RBF kernels may not be optimal in certain situations, such as when the

number of parameters is large [44]. One of the most widely used RBF kernels is the Gaussian kernel [10].

Gaussian RBF Kernel:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$$

SVMs are one of the most popular classification algorithms because they are well suited to deal with data with high dimensionality. Much research in the field of computer includes high dimensionality data. An in-depth discussion of SVMs can be found in [15].

FACS AU-CODED IMAGE DATABASES

Cohn-Kanade

The Cohn-Kanade AU-Coded Facial Expression Database (CK) was compiled in an effort to advance the field of facial expression analysis [45]. The CK database contains a substantial number of images with a range of expressions and uses FACS to define those expressions. Certified FACS coders were used to code sequences in the database. 201 participants ranging from 18 to 50 years old, with an ethnic distribution of 81% Euro-American, 13% African-American, and 6% other, took part in the original data collection process. 69% of the subjects were female.

Data was collected for facial expressions of 201 participants using two hardware synchronized Panasonic WV-3230 cameras. One camera was front-facing, and the other was 30° to the right for each subject. Ambient room lighting coupled with a high-intensity lamp was used for one-third of the subjects. Two-high intensity lamps, using reflective umbrellas to attain homogeneous lighting conditions, were used for two-thirds of the subjects. Frontal-view sequences were digitized into 8-bit grayscale or 24-bit color with resolutions of either 640x490 or 640x480.



Figure 9: Example of an image from the CK+ Database [54]

The initial release of the CK database contained 486 FACS-coded sequences across 97 subjects. A sequence included all frames from the neutral expression to the peak expression. Each peak expression frame was FACS AU-coded. Emotion labels were added, but not validated. Version 2, also known as CK+, added emotion labels. The 97 subjects were university students ranging from 18 to 30 years old. 65% of the participants were female. Each participant was asked to perform 23 different facial expressions. These expressions included single action unit movements, such as AU 12 for lip corner puller, and also combinations of action unit movements, such as AU 1 + 2 for inner and outer brow raiser. Full details about the data collection

process for the full database can be found in [45].

Although the Cohn-Kanade database is one of the most used face databases in the world, facial expression recognition and related areas of research have advanced substantially since the original release of the database. As a result, three main limitations of the CK database became apparent. The first limitation was that the original emotion labels were not properly validated. The expression labels were those requested of the subject, not the actual expression performed. The second shortcoming was the lack of a common performance metric. It was difficult for researchers to evaluate new algorithms with no existing benchmark with which to compare. The third limitation was that to make quantitative meta-analysis possible a standard protocol for common databases was needed. These three issues were addressed in the release of Version 2, also known as the Extended Cohn-Kanade Database or CK+ [54].

CK+ added 107 new sequences from 26 new subjects, bringing the total for the database to 593 sequences across 123 subjects. Action Unit labels and emotion labels have been revised and validated. Non-posed smiles and related meta-data have also been added. Full details about the CK+ database and related studies can be found in [54].

Bosphorus

The Bosphorus database, developed at B gaziçi University in Istanbul, Turkey, is a FACS AU-coded 3D facial database consisting of various facial expressions, facial occlusions, and head poses [64]. Corresponding 2D high-resolution color images are also available. 20 lower face action units, 5 upper face, and 3 combinations of FACS action units were used for the database, and for Version 2, six prototypical emotions of happiness, sadness, anger, fear, disgust, and surprise were captured.

The current database consists of 4666 face images in a variety of poses, expressions, and occlusion conditions. Occlusions include facial hair, moustaches, glasses, etc. Of the 105 subjects, 60 are male, and the majority are Caucasian and between 25 and 35 years old. Each image has been manually labeled with 24 landmark points covering the eyebrow, eye, nose, mouth, and ear lobe regions.

Data collection was performed using an InSpeck Mega Capturor II 3D digitizer device with an image resolution of 1600x1200 pixels and sensor dimensions of 0.3mm, 0.3mm, and 0.4mm (x,y,z). Subjects were placed approximately 1.5 meters from the capturing device and a 1000w halogen lamp was used for lighting. Lighting conditions were similar for each subject and as a result lighting variances across the database are minimal. Due to the



Figure 10: Example of an image from the Bosphorus Database [65]

system used, capturing video of the subjects was not possible. Originally, images were FACS labeled based on the requested expression. The current database has been AU-coded by a certified FACS coder [65].

The original database had two versions. Version 1 contained 34 subjects with approximately 10 expressions, 13 poses, 4 occlusions, and 4 neutral faces per subject. Version 2 added 47 subjects with approximately 34 expressions, 13 poses, 4 occlusions, and 1 or 2 neutral faces per subject. Version 2 included the six prototypical emotion expressions for a majority of the subjects. Examples of the desired facial expressions were shown to the subjects so they could be mimicked. When collecting emotional expressions, subjects were

not shown any examples. In an attempt to enhance expression realism, 29 participants were professional actors/actresses. All expressions were posed, and not every expression is available for each subject, as some subjects could not properly produce some of the desired action units. In every condition but “hair occlusion”, subjects wore headbands to keep hair above their forehead.

Three types of head poses were captured for each individual resulting in 7 yaw angles, 4 pitch angles, and 2 cross rotations (both pitch and yaw). The action chosen for eye and mouth region occlusions was left to each subject and as a result, varied in both the action performed and the amount of facial occlusion. A collection of eyeglasses were available for the subjects to use during the eyeglasses occlusion portion. If subjects’ hair was long enough hair, hair occlusions covering part of the face were included.

PREVIOUS WORKS

Automatic Extraction of Action Units

Bartlett et al. used feature extraction techniques such as Gabor filters, Support Vector Machines (SVMs), and Hidden Markov Models (HMMs) to automatically FACS AU-code facial expressions. In one study they attempted to automatically code 12 (6 upper face and 6 lower face) action units for posed facial expressions for 20 subjects. Gabor Wavelet decomposition and independent component analysis performed best with a 95.5% accuracy rate over five other approaches, which were optical flow (85.6%), local feature analysis (81.1%), and eigenfaces (79.3%), Fisher's linear discriminant (75.7%), and explicit features (wrinkles) (57.1%). Another study involved 17 subjects in a high-stake mock crime experiment conducted by Mark Frank and Paul Ekman. To address out-of-plane head rotation, which is a common element of spontaneous expressions, they used deformable 3D models to warp a subject's face to a canonical face geometry. Blinks (AU 45), brow raiser (AU 1 + 2), and brow lowerer (AU 4) were chosen as the three actions to recognize. Best results were achieved using HMMs trained on the outputs the SVM for Blink versus Non-Blink actions, with an accuracy rate of 98.2% [8, 13].

In a later study by Bartlett et al., a real-time automatic classification system for both posed and spontaneous facial expressions was evaluated. The image databases used were RU-FACS from Rutgers University, the Cohn-Kanade database, and a database of images collected by Ekman and Hager (Ekman-Hager Database). RU-FACS contains FACS AU-coded spontaneous facial expressions from 100 subjects. The Cohn-Kanade and Ekman-Hager databases both contain FACS AU-coded posed facial expressions, and for this study 119 subjects were included. The technique that provided the best results, in both speed and accuracy, was the method of selecting a subset of Gabor filters using Adaptive Boosting (Adaboost) and using those outputs to train SVMs. Automatic FACS labeling for spontaneous expressions and posed expressions achieved 90.5% and 94.8% agreement with human FACS codes, respectively. Recognition of full facial expressions of emotion in a 7-way forced choice achieved 93.3% accuracy. [5, 6].

The Computer Expression Recognition Toolbox (CERT) is a real-time, fully automated software tool that uses appearance-based discriminative approaches, such as Gabor filters and SVMs, to code real-time video with respect to “40 continuous dimensions, including basic expressions of anger, disgust, fear, joy, sadness, contempt, a continuous measure of head pose (yaw, pitch, and roll), as well as 30 facial action units (AU’s) from the Facial Action Coding System” [50, 3, 4]. Although CERT is a relatively new tool it

has already been used in a variety of facial expression recognition studies including drowsiness detection, affect-sensitive adaptive tutoring, and games for children with *Autism Spectrum Disorder* (ASD) [78, 16, 80, 49, 19].

CERT was used in a study by Littlewort et al. to distinguish real versus faked expressions of pain. 48 subjects took part in an experiment which used cold pressor pain to induce real expressions of pain [51]. A subject's forearm was submerged in 5°C ice water to induce real pain expressions. The water was 20°C for baseline and faked pain conditions. For the faked pain portion of the experiment, subjects were asked to produce facial expressions that would convince an expert they were genuinely experiencing pain. To test human accuracy in differentiating real from faked expression of pain, 170 naïve observers viewed videos of the experiment and were asked to choose whether the subject's pain expression was real or faked. Naïve observer mean accuracy was 52%. The automatic system used a two-stage process to differentiate real from faked pain. It first automatically detected facial actions and then provided that data to a machine learning classifier. A few meaningful differences in action unit activity and intensity for real versus faked pain were observed, including more brow lowerer (AU 4), cheek raiser (AU 6), and inner brow raiser (AU 1) activity during faked pain conditions. The automatic system achieved .72 area under the ROC curve, which is equivalent to 72% accuracy on a 2-alternative forced choice of fake versus real pain. In a later study by

the same authors naïve human observer mean accuracy was 49.1% with a standard deviation of 13.7%. The automatic system achieved 88% accuracy [52].

Using 364 sequences across 94 subjects, from the Cohn-Kanade database, Gonzalez et al. developed a system to detect individual lower facial action units using a geometry-based approach [42]. The shape model used consisted of 83 facial feature points. AdaBoost and overlapping coefficients (OVL) were used for feature extraction and AdaBoost and SVMs were used for classification (AdaBoost is both a feature selection and classification technique). An average accuracy of 94.55% was achieved.

Valstar et al. examined spontaneous versus posed facial expressions by developing a geometry-based system capable of automatic analysis of brow actions [76]. To automatically distinguish posed from spontaneous brow actions they focused on the temporal dynamics of the brow actions. Temporal dynamics refers to the neutral, onset, apex, and offset stage of facial actions. Brow actions were chosen because the brow is active in a variety of facial expressions. A shape model consisting of 8 feature points representing the eyebrow, eye, and nostril regions of the face was used in the experiment. FACS action units 1, 2, and 4 were represented by the locations of these points. The MMI Facial database, Cohn-Kanade Facial database and DS118 dataset were used in this study [57, 74, 62]. The MMI Facial Expression

and Cohn-Kanade Facial Expression databases consist of 123 posed facial expressions and the DS118 dataset consists of 139 samples of spontaneous facial expressions. Gentle Boost and SVMs were used to detect AU activity and Relevance Vector Machines (RVM) was used to classify the data. A correct classification rate of 90.7% was achieved during testing across the three datasets.

A similar experiment was conducted by Sebe et al. using a variety of classification techniques such as Bayesian networks, decision trees, SVMs, and k-Nearest Neighbor for affect recognition [68].

Using AAMs for Expression Recognition

Benedikt et al. explored the idea of using facial emotions as a behavioral biometric using the Facial Action Coding System (FACS) for expression definition and Active Appearance Models (AAMs) for feature extraction [11]. 3-D video data was collected over an extended time interval for both verbal and non-verbal facial actions. The experiment examined identification and verifications problems and results showed that while nonverbal facial actions (emotional expressions) were not sufficiently reliable for identity recognition, verbal facial actions (speech-related) showed potential for future use in biometric applications.

Patterson et al. also explored the idea of using facial actions as a biometric. They choose to use a blink as the facial action, AAMs for feature extraction, and HMMs for classification. 23 landmark points marked the eyebrow, eye socket, and eyelid regions. Video data resulting in 1100 image sequences across 12 individuals was collected. They achieved 100% accuracy for all tests [58].

Ratliff explored the use of AAMs and three popular classification techniques for recognizing facial emotional expressions. The classification techniques he employed include Euclidean distance, Gaussian Mixture Models (GMMs), and SVMs and the AAM landmark scheme used 113 points. The experiment used sequences from 18 subjects displaying emotions of fear, joy, surprise, anger, disgust, and sadness from the Face and Gesture Recognition Research Network (FG-Net) database. A neutral expression was also included. Euclidean distance measure, Gaussian Mixture Models (GMMs), and SVMs achieved a mean accuracy of 83.9%, 87.1%, and 91.3%, respectively [60].

Lucey et al. performed experiments in an attempt to recognize facial action units for posed and spontaneous facial expressions [55]. The Cohn-Kanade dataset was used for posed expressions and the RU-FACS dataset was used for spontaneous expressions. The experiment used AAMs for feature extraction and SVMs for classification. The use of a Nearest Neighbor

(NN) classifier based on either PCA or LDA subspaces was considered, but preliminary tests showed it offered no advantage over the SVM approach. Results for posed action unit recognition showed an improvement over previous research. Most notably, this study’s approach of using active appearance models outperformed the approach of using Gabor filters with AdaBoost from Bartlett et al. [7], when using the same database and set of action units. Results from the spontaneous expression experiment were poor, mainly due to the problems resulting from the amount of head movement in subjects. Complete details, including statistical analysis of each experiment, can be found in [55].

Ashraf et al. used the UNBC-McMaster shoulder pain expression archive in a study using AAMs for feature extraction and SVMs for classification, to differentiate real from faked expressions of pain [2]. The study found that decoupling the face into separate non-rigid shape and appearance components increased performance for their study. Unlike Littlewort et al., they attempted to detect pain expressions not at the action unit level but rather from shape and appearance features. Many previous approaches applied techniques that used rigid registration, however, they claim, based on their findings, that rigid registration of appearance may not be necessary for some applications.

Lucey et al. added to the Cohn-Kanade Database in order to address

a few limitations [54]. One of these limitations was the lack of a common performance metric, or benchmark, to which other researchers could compare their own algorithms. Therefore, the authors conducted their own experiment using the new data. They used Active Appearance Models to track faces and extract features. A linear one-vs-all, two-class SVM was used to detect the presence of each AU. 17 AU's were included in this study along with seven stereotypical emotions, as defined by FACS. These emotions are anger, contempt, disgust, fear, happiness, sadness, and surprise.

The authors used a 3-step process for selecting the appropriate emotion label for each peak frame in a sequence. The first step took a strict approach by evaluating a sequence by the action units present, based on the Emotion Prediction Table from the FACS manual [34]. The Emotion Prediction Table lists the action units required for each prototypical emotion and the major variants of each emotion. If a sequence satisfied these requirements for any emotion, it was provisionally given that emotion label. The first step in the selection process was considered "strict" because the presence of, or lack of, any action unit not included in the table would result in the clip being excluded. The second step took a more loose approach by allowing action units not included in the prototypical emotions, or major variants of the emotions, if they were consistent with the emotion displayed. For instance, AU 4 is consistent with negative emotions such as anger, but not with positive emo-

tions such as happiness. The third step was a visual evaluation of a sequence to determine if the expression resembled the targeted emotion category. 327 of the 593 sequences met criteria for one of the seven prototypical emotions. Complete details about the selection process can be found in [54].

AU detection and emotion detection were the two experiments conducted on the CK+ database. Leave-one-subject-out-cross-validation resulted in 123 different training and testing sets for AU detection and 118 different training and testing sets for emotion detection.

Similarity-normalized shape (SPTS) and canonical-normalized appearance (CAPP) were derived once the shape and appearance AAM parameters were computed. SPTS refers to the shape information of an image and CAPP refers to the texture information. SPTS performed well for emotions that caused distinct changes in the shape of the AAM mesh, such as disgust and happiness. Conversely, CAPP performed well for emotions that caused texture changes, such as anger and sadness. The combination of features (SPTS + CAPP) produced the best recognition rates for all emotions except surprise. These recognition rates were Anger 75%, Disgust 94.7%, Fear 65.2%, Happiness 100%, Sadness 68%, and Contempt 84.4%. SPTS achieved a 100% recognition rate for surprise.

Cohn et al. developed an automated system to analyze facial displays [21]. 100 subjects performed facial actions on video and the resulting image sequences were hand coded by certified FACS coders. The automated system used feature point tracking to determine facial movements for individual, and a combination, of 15 action units. Feature point measurements were captured for the horizontal and vertical displacements of 8 feature points located in each eye region. Linear discriminate analysis was performed on the feature point measurements. Among other analysis, the researchers compared the recognition accuracy for non-Duchenne (AU 12) versus Duchenne smiles (AU 6 + 12). Average accuracy for non-Duchenne versus Duchenne smiles was 83% for the training set and 82% for the cross-validation set. In the cross-validation set, average agreement with human FACS codes for the brow, eye, and mouth regions was 91%, 88%, and 81%, respectively.

Cohn et al. studied differences in timing of smile onsets of spontaneous and deliberate smiles [20]. In a similar study, Delannoy et al. used temporal-based techniques to model the aspects of posed and spontaneous facial expressions [30].

Littlewort et al. attempted to discriminate Duchenne from non-Duchenne smiles by focusing on the eye region [53]. One dataset was collected by C. Ri-

ley of the BBC and contained images of both genuine and posed smiles (BBC set). Human observers were shown images from the BBC set and asked to decide if the subject was displaying a genuine or posed smile. The mean accuracy rate for human observers was 60%. There were 40 images in the test set and the researchers system achieved 75% accuracy using a linear SVM on normalized Gabor filter outputs. A combined dataset from the University of Pittsburgh and University of California, San Francisco (Ekman-Pittsburgh set) contained 157 Duchenne and 72 non-Duchenne smiles. Using Gabor filters and a linear combination of SVMs resulted in the best accuracy rate, which was 87%. The experiment results reflected two significant factors related to performance. The first is that linear SVM kernels did not perform as well as polynomial or Gaussian kernels. The second was that Gabor filters contributed to the overall classification performance by enhancing SVM performance.

Valstar et al., used a geometric-based approach to distinguish posed smiles from spontaneous smiles [75]. The study focused on three main issues: the role of the head and shoulders in the recognition of posed versus spontaneous smiles, the temporal dynamics associated with non-verbal behavior in regards to posed versus spontaneous smile recognition, and the different multi-modal data fusion strategies and how they compare to monomodal classification results. Three fusion strategies were examined: early, mid-level, and late fusion.

Low-abstraction features are used for early fusion, while high-abstraction features are used for mid-level and late-level fusion. A cylindrical head tracker was used to track head motion and two particle filtering techniques were used to track facial actions and shoulder movements. 12 points were used for tracking activity in the face, and 5 points were used for tracking movement in the shoulder region. 100 videos displaying posed smiles and 102 videos displaying spontaneous smiles were used from the MMI facial expression database [57, 74]. Using Gentle Boost and SVM, 94% of videos were classified correctly. The late fusion strategy performed the best. The researchers concluded that late fusion performed the best because it allows for the problem to be decomposed into smaller sub-problems and also because it uses high-abstraction features. The former allows for the training of specialized classifiers and the latter encodes important temporal dynamic attributes associated with nonverbal behavior.

EXPERIMENT

Overview

The system described in this paper uses active appearance models for feature extraction because AAMs have been shown, in other studies, to be a good feature extraction technique. The three machine learning techniques used were chosen because of their common use in affect recognition studies. While there have been studies that have differentiated Duchenne from non-Duchenne smiles relatively successfully, their methodologies differ from this study, in that none of them use active appearance models for feature extraction. To the best of the author's knowledge, no study exists currently that differentiates Duchenne from non-Duchenne smiles using active appearance models. It's the author's belief that the use of AAMs for feature extraction, and three popular machine learning classifiers chosen for this study will result in favorable or better performance than other published methodologies for similar studies.

The presence of the Duchenne marker (orbicularis oculi, pars lateralis) is what distinguishes a Duchenne smile from a non-Duchenne smile. Very few individuals have the ability to voluntarily contract the orbicularis oculi, pars

lateralis muscle, and studies have shown, based on brain activity and other observations, that the presence of the Duchenne marker signifies genuine smiles [38, 36, 35]. A system that can differentiate the subtle differences between Duchenne from non-Duchenne smiles is useful for a multitude of applications in a variety of fields, such as affect analysis in clinical psychology. If successful, the methodology proposed in this study could be used in other affect recognition systems, such as those used for deception detection.

Data Collection

The Extended Cohn-Kanade AU-coded Facial Expression database and Bosphorus FACS AU-coded database were used for this experiment. 104 images from the CK+ database and 135 images from the Bosphorus database were used in the experiment, resulting in a total of 239 images across 182 subjects. Of the 104 images from the CK+ database, 69 were Duchenne and 35 were non-Duchenne. Of the 135 2D images from the Bosphorus database, 47 were Duchenne and 88 were non-Duchenne. This resulted in a total of 116 Duchenne smile and 123 non-Duchenne smile images. Image selection was a 3 step process.

The first step examined each image for the presence of AU 6 (Cheek Raiser - Orbicularis oculi, pars lateralis) and AU 12 (Lip Corner Puller - Zygomatic

major). If AU 12 was present, the image was checked for AU-codes inconsistent with Duchenne and non-Duchenne smiles, such as AU 4 (Brow Lowerer). Other action units inconsequential to Duchenne and non-Duchenne expressions were allowed, such as AU 14 (Dimpler).

Once this step was complete, the remaining images were sorted into the Duchenne group if they contained AU 6 and the non-Duchenne group if AU 6 was not present. From here, images in the Duchenne group from the Bosphorus database with A (Trace) intensity were discarded. The Bosphorus database has been FACS coded by a single certified FACS coder [65]. Typically, as in the case of the Cohn-Kanade database, at least two independent FACS coders are used [45]. Upon examination of the images in the Duchenne group the author had concerns with the accuracy of AU 6A scored images from the Bosphorus database, but not for any AU 6 scored images from CK+ database. Therefore, the author made the decision to exclude Bosphorus images that contained AU 6A.

The final step of the selection process involved examining the remaining images and determining if they fit the targeted facial expression, contained any occlusions, or were outliers. This step was not independent of the first two steps, as determining which AUs to allow and which to exclude had to be examined on a per AU basis. There were 4 outliers (2 subjects) due to ethnicity. They were removed from the final dataset.

The selection process for this study was based heavily on the selection process described in the validation study of the Cohn-Kanade Extended database by Lucey et al., as described in [54].

Methodology

116 Duchenne and 123 non-Duchenne smile images composed the experiment dataset. A Constrained Local Model (CLM) was used to auto-landmark the images using a 227-point landmark scheme [27, 63]. A manual examination of each landmarked image was performed and corrections were applied when necessary using a custom landmark editor. A 227-landmark scheme was chosen because it was readily available for use by the author and it adequately covers facial features that have been shown to be important for distinguishing Duchenne from non-Duchenne smiles [33, 41].

An Active Appearance Model was built using all 239 images, retaining the top 95% eigenvectors for shape, top 95% eigenvectors for texture, and the top 95% eigenvectors for the combined appearance. Using the resulting AAM model and the 239 images, a fitting process was performed, resulting in 239 feature vectors with each containing 55 feature weights (Referred to in the results section as *100%-100%*). Using all of the images to build the AAM model allows the feature vectors returned to accurately statistically

represent, with respect to the model, the original images and, thus, is known as the *ideal-case* approach. To determine how well AAMs generalize to real-world data, a second model was created using roughly 50% of the images (120 images), with each containing 41 feature weights, so that no faces used to construct the AAM model were used in classification training or testing. For a *semi-generalized* approach, 100% of the feature vectors were used as input to the classifiers (50%-100%). For a *fully-generalized* approach, only the feature vectors from the images not used to create the model, were used as input to the classifiers (50%-50%).



Figure 11: The non-Duchenne, Duchenne, and combined (non-Duchenne and Duchenne) average face.

The feature vectors were used as input to k -Nearest Neighbor, Naïve Bayes, and Support Vector Machine classifiers, using the open-source, machine learning software suite *Orange* [28]. *Orange* was chosen because it is a widely-used software suite that contains a variety of machine learning classifiers and, given

its open-source nature, is highly customizable. For the k -Nearest Neighbor classifier, a $k = 5$ and Euclidean distance metric were chosen. Naïve Bayes used the default *Orange* parameter settings of *Relative Frequency* as the probability estimator for prior class probabilities and a LOESS window size of .05 and 100 LOESS sample points. *Orange* uses *LIBSVM* which is one of the most popular and widely used SVM libraries available [17]. The choice to use a ν -SVM, Gaussian RBF kernel in this study was based on findings and suggestions that ν -SVMs are preferred over C -SVMs, and RBF kernels are preferred over other kernels for classifying nonlinear data [67, 44, 18, 10, 17]. The data was normalized prior to learning and a grid-search, as mentioned in Section 3.4 [44], was performed using values ranging from 0.01 to 1, at .01 increments, to determine optimal ν and γ values, which were .03 and .5, respectively.

Leave-one-subject-out cross-validation was performed on the data. This functionality is not currently available using the pre-existing *Orange* methods. A *leave-one-subject-out* method was written, implementing other *Orange* methods. To confirm its accuracy, each vector was labeled as a unique subject, and it was tested against *Orange's* *leave-one-out* algorithm.

Three approaches were compared: ideal-case (*100%-100%*), semi-generalized (*50%-100%*), and fully-generalized (*50%-50%*). For each approach, Area Under the ROC Curve (AUC) scores, as well as a confusion matrix, were reported

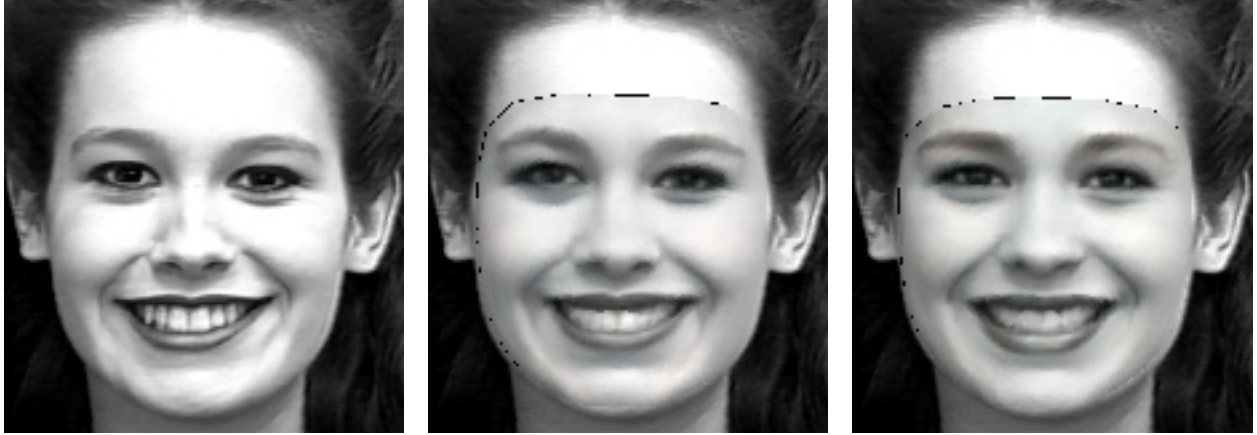


Figure 12: The original image, the reconstructed image from the *ideal-case* approach, and the reconstructed image from the *fully-generalized* approach. The quality of the reconstruction is high, which shows that the AAM, with both approaches, is accurate statistically in representing the original image.

for each of the three classifiers. AUC was chosen as the performance metric because it has been shown to be more reliable and contain more preferable properties than classification accuracy [14, 6, 48]. The confusion matrices contain the predicted class and actual class for Duchenne and non-Duchenne smiles for each approach for a specific classifier. They show the distribution of correct and incorrect classifications and this may give insight to whether or not the classifier is suffering from the problem of over-fitting. The confusion matrices can be found in Tables 3-11 which are located in *Appendix A*.

RESULTS AND DISCUSSION

Results

For the *ideal-case*, naïve Bayes performed the worst at 88.23%, k-Nearest Neighbor (kNN) was slightly better at 91.81%, and Support Vector Machine (SVM) was, by far, the best performing classifier at 99.00%. The *semi-generalized* approach saw a slight increase in performance of naïve Bayes at 88.46%, and slightly poorer performance of kNN and SVM, at 91.65% and 97.92%, respectively. The *fully-generalized* approach had the worst performance of the three, with naïve Bayes at 83.22%, kNN at 91.60%, and SVM at 95.83%. Table 1 contains the performance of the three classifiers for each generalization approach. Figure 13 contains a visual comparison of the Area Under the ROC Curve (AUC) scores for the three classifiers for each approach.

Method	Ideal-Case	Semi-Generalized	Fully-Generalized
kNN	91.81%	91.65%	91.60%
Naïve Bayes	88.23%	88.46%	83.22%
SVM	99.00%	97.92%	95.83%

Table 1: Area under the ROC Curve Score (2-Alternative Forced Choice)

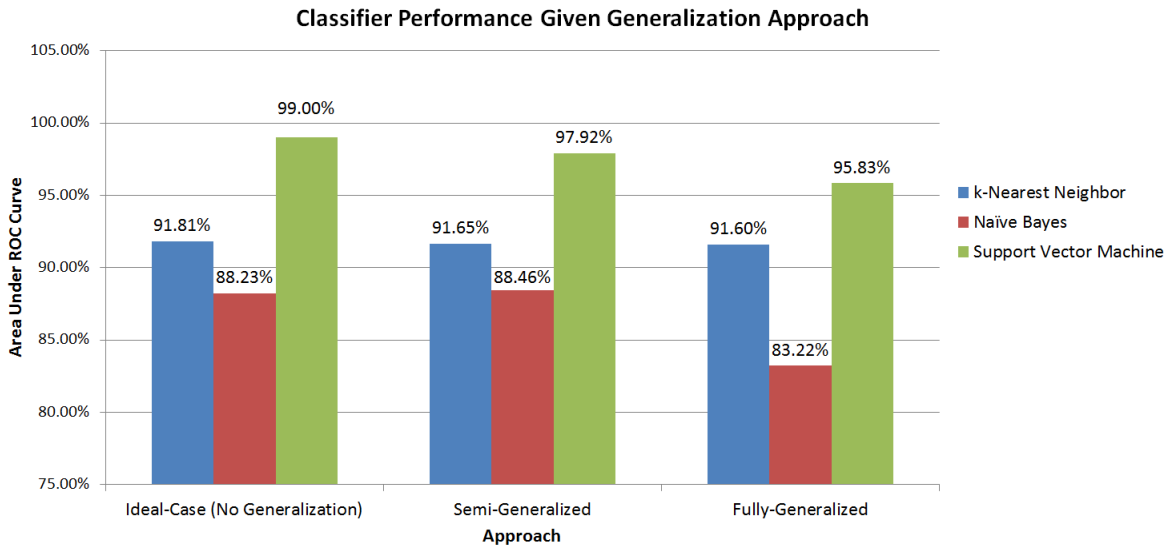


Figure 13: Classifier performance given generalization approach

Discussion

The three hypotheses were developed to evaluate three main points. Hypothesis I was created as a way to determine if the proposed system would perform better than a “random classifier”, which is a common technique used to determine if a system would outperform a naïve human observer. All three approaches performed significantly better than a “random-chance” classifier.

Hypothesis II used the *ideal-case* approach. Since the images used to build the AAM model were the same images used for getting feature vec-

tors, the returned feature vectors were statistically accurate, with respect to the AAM model. This hypothesis was created to determine if the features extracted contained any discerning information in regards to Duchenne and non-Duchenne smiles, and if they did, how well the classifiers could differentiate between the two facial emotion expressions. SVMs outperformed previously published methodologies, while kNN and naïve Bayes compared favorably.

Study	Image Database	Feature Extraction	Classifier	Score
Littlewort et al. [53] ¹	BBC	Gabor Filters	SVM	75.00%
Cohn et al. [21] ¹	Cohn-Kanade	Optical Flow (Lucas-Kanade)	Linear Discriminant Analysis	82.00%
Littlewort et al. [53] ¹	Ekman-Hager, Pittsburgh	Gabor Filters	SVM	87.00%
Cohn et al. [20] ²	Cohn-Kanade	Tian, Kanade, and Cohn System [71]	Linear Discriminant Analysis	93.00%
Valstar et al. [75] ²	MMI Facial Expression Database	GentleBoost	SVM	94.00%
<i>This Study</i>	Extended Cohn-Kanade (CK+), Bosphorus	AAM	SVM	95.83%

Table 2: Comparison of similar published approaches. ¹ denotes Duchenne vs non-Duchenne. ² denotes posed vs. spontaneous study. (Note: If a study used multiple techniques, only the best performing combination and performance score is listed)

Assuming the results support hypothesis II, which would support the belief that there is sufficient discerning information about the differences between Duchenne and non-Duchenne smiles in the features extracted using AAMs,

hypothesis III was crafted for evaluating how well AAMs generalize previously unseen Duchenne and non-Duchenne images. That is, whether or not the returned feature vectors are statistically accurate representations of the original images, with respect to the AAM model. With this approach, a fair comparison to published methodologies from similar studies could occur, and the results would also give insight to how the methodology described in this study may perform in *real-world* applications, in which the subject would most likely not be part of the AAM model. The combination of AAMs and SVMs resulted in performance better than previously published methodologies, while kNN and naïve Bayes compared favorably.

The confusion matrices contain the predicted class and actual class for Duchenne and non-Duchenne smiles for each approach for a specific classifier. The confusion matrices can be found in Tables 3-11 which are located in *Appendix A*. The common pattern observed for all three approaches is that for the kNN classifier, more Duchenne smiles are incorrectly classified as non-Duchenne smiles; and for naïve Bayes and SVM classifiers, more non-Duchenne smiles are incorrectly classified as Duchenne smiles. These misclassifications are most likely due to the natural bunching of skin around the eye region, which can occur due to a number of factors, such as aging and habitual facial expressions. This bunching of skin can sometimes be misclassified as the contraction of the orbicularis oculi, pars lateralis even by expert human

FACS AU-coders. These misclassification results are not worrying because of the relatively low percentage of misclassified smiles and the consistency of the misclassification across the three approaches.

All three hypotheses were supported by the results of this study and the performance of the techniques used in this study compared favorably or performed better than other published methodologies for similar studies. Similar studies include not only differentiating Duchenne from non-Duchenne smiles, but also differentiating posed from spontaneous (genuine) smiles. The majority of posed versus spontaneous smile studies focus on the temporal dynamics (*i.e. onset-apex-offset*) of smiles and the geometric features, rather than the shape/texture/combined appearance information used to distinguish one emotion from another. However, they do attempt to distinguish posed from genuine smiles, which is similar to this study. In the majority of cases *Duchenne vs. non-Duchenne* and *genuine vs. posed* refer to the same types of smiles [38, 36, 35]. For this reason, a few applicable posed vs. spontaneous study results have been included in the comparison of results, which can be found in Table 2.

The technique of using AAMs for feature extraction and SVMs for classification results in, to the best of the author’s knowledge, the best performance for differentiating Duchenne from non-Duchenne smiles. The robustness of SVMs makes it an excellent choice for classification, and, given the results

and other related published research, it would appear AAMs is an excellent feature extraction technique. It is the author's belief that using AAM for feature extraction is the main reason this study results were either favorable or better than the results of previously published methodologies. Many of these similar studies also found SVMs to be the best classification technique, and because they used the same databases, or similar FACS AU-coded databases, the main difference between the other published methodologies and this study is the chosen feature extraction technique.

FUTURE WORK

There are a few modifications that could be implemented to further examine the quality of using AAMs for feature extraction and kNN, naïve Bayes, and SVMs for classification of Duchenne versus non-Duchenne smiles. Adding more images from FACS AU-coded databases, with increased ethnic diversity would help to test real-world performance. While the system in this study has been shown to effectively differentiate Duchenne from non-Duchenne smiles, the system used databases where the majority of subjects are Caucasian. Using an ethnically diverse image database, the system proposed in this study may be less effective, as the ability to discern between the two subtle facial emotions may be less effective for certain ethnicities or given a diverse range of ethnicities in the AAM model. This study performed a 2-alternative forced choice, for Duchenne and non-Duchenne smiles. Another modification that would help to evaluate the methodology described in this paper, for differentiating Duchenne from non-Duchenne smiles, is to use a dataset with other FACS expert AU-coded emotions. Real-world applications will require the ability to recognize a variety of emotions, so it is important that a system can first differentiate smile-based emotions, such as non-Duchenne, Duchenne, miserable, and concealment smiles, from other emotions, such as anger or surprise, before differentiating the subtle emotion expression differ-

ences in Duchenne versus non-Duchenne smiles.

While a 227-point landmark scheme worked well for this study, it would be interesting to observe how performance changes, given a different scheme. The 227-point scheme seems to cover the appropriate features of a face for differentiating Duchenne from non-Duchenne smiles, but it is possible that a different landmark scheme may result in better performance. Only further experiments could show if this is the case. The modifications listed in the previous paragraph may also be affected by the landmark scheme choice. It would be interesting to see if the landmark scheme chosen has a significant effect on the performance of such systems, and is the reason that this is a popular, active area of research.

One of the main challenges for affect recognition research is acquiring large datasets of FACS AU-coded images. It is a time consuming and expensive process to capture facial emotional expressions and although there are a substantial number of emotion labeled datasets publicly available, many of them contain custom, complex organization schemes [40]. To help solve this problem the MMI Facial Expression Database project has developed and is constantly expanding a freely available, online, searchable resource, which “aims to deliver large volumes of visual data of facial expressions to the facial expression analysis community” [57, 74]. The MMI Facial Expression database could prove to be a powerful and useful tool for many emotion

recognition researchers if it achieves its goal of being a centralized source for FACS expert AU-coded facial emotion expression data.

Properly analyzing facial emotion expression is critical for the success of a variety of real-world applications in a variety of fields, such as psychology, medicine, and law enforcement. A psychologist could determine if the emotion being expressed by a patient is consistent with their facial emotion expressions. This may be an important step for accurately classifying certain “negative emotions” that, for a variety of reasons, a patient may try to conceal. There are also many uses in telemedicine applications, where a doctor or nurse is not physically present and must rely on the honest and accurate self-evaluation of a patient. Whether consciously or not, the patient may exhibit useful facial expression information, which is useful for healthcare professionals, as supported recently in a study by Schmidt et al. [66]. For law enforcement, a system used for deception detection may be a useful tool for determining the validity of the information a suspect is providing.

This study evaluates a methodology that, to the best of the author’s knowledge, has not yet been used to differentiate Duchenne from non-Duchenne smiles, commonly referred to as genuine vs. posed. Accurately differentiating between these two facial emotion expressions is applicable for the three applications mentioned for psychology, medicine, and law enforcement. The system described in this paper must be modified if it is to be used in real-

world applications, with the most notable modifications being expanding the ethnic diversity and types of facial emotion expressions. However, the strong performance of this methodology is a positive step towards successful affect recognition and deception detection systems.

REFERENCES

- [1] Zara Ambadar, Jeffrey Cohn, and Lawrence 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. 10.1007/s10919-008-0059-5.
- [2] Ahmed Bilal Ashraf, Simon Lucey, Jeffrey F. Cohn, Tsuhan Chen, Zara Ambadar, Kenneth M. Prkachin, and Patricia E. Solomon. The Painful Face - Pain expression recognition using active appearance models. *Image Vision Comput.*, 27:1788–1796, November 2009.
- [3] M. Bartlett, G. Littlewort, J. Whitehill, E. Vural, T. Wu, K. Lee, A. Ercil, M. Cetin, and J. Movellan. Insights on spontaneous facial expressions from automatic expression measurement. In M. Giese, C. Curio, and H. Bulthoff, editors, *Dynamic Faces: Insights from Experiments and Computation*, chapter 14. MIT Press, 2010.
- [4] M. Bartlett and J. Whitehill. Automated Facial Expression Measurement: Recent applications to basic research in human behavior, learning, and education. In James V. Haxby Andrew Calder, Gillian Rhodes and Mark H. Johnson, editors, *Handbook of Face Perception*. Oxford University Press, 2011.

- [5] Marian Stewart Bartlett, Gwen Littlewort, Mark Frank, Claudia Lainscsek, Ian Fasel, and Javier Movellan. Recognizing Facial Expression: Machine learning and application to spontaneous behavior. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:568–573, 2005.
- [6] Marian Stewart Bartlett, Gwen Littlewort, Mark Frank, Claudia Lainscsek1, Ian Fasel, and Javier Movellan. Fully automatic facial action recognition in spontaneous behavior. *Automatic Face and Gesture Recognition, IEEE International Conference on*, 0:223–230, 2006.
- [7] Marian Stewart Bartlett, Gwen Littlewort, Claudia Lainscsek, Ian Fasel, and Javier Movellan. Machine learning methods for fully automatic recognition of facial expressions and facial actions. In *Proc. IEEE Intl Conf. Systems, Man and Cybernetics*, pages 592–597, 2004.
- [8] Marian Stewart Bartlett, Javier R Movellan, Gwen Littlewort, Bjorn Braathen, Mark G Frank, and Terrence J Sejnowski. Towards automatic recognition of spontaneous facial actions. *What the face reveals Basic and applied studies of spontaneous expression using the facial action coding system FACS Second Edition*, (1990):393–426, 2001.
- [9] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. of the Royal Soc. of London*, 53:370–418, 1763.

- [10] Asa Ben-Hur and Jason Weston. A user's guide to support vector machines. *Methods in Molecular Biology. Part 2*, 609:223–239, January 2010.
- [11] Lanthao Benedikt, Darren Cosker, Paul L. Rosin, and David Marshall. Assessing the uniqueness and permanence of facial actions for use in biometric applications. *Trans. Sys. Man Cyber. Part A*, 40:449–460, May 2010.
- [12] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory, COLT '92*, pages 144–152, New York, NY, USA, 1992. ACM.
- [13] Bjorn Braathen, Marian Stewart Bartlett, Gwen Littlewort, Evan Smith, and Javier R. Movellan. An approach to automatic recognition of spontaneous facial actions. *Automatic Face and Gesture Recognition, IEEE International Conference on*, 0:0360, 2002.
- [14] Andrew P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn.*, 30(7):1145–1159, July 1997.
- [15] Christopher J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2:121–167, June 1998.

- [16] Nicholas J. Butko, Georgios Theodorou, Matthai Philipose, and Javier R. Movellan. Automated facial affect analysis for one-on-one tutoring applications. In *FG*, pages 382–387. IEEE, 2011.
- [17] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2:27:1–27:27, May 2011.
- [18] Pai-Hsuen Chen, Chih-Jen Lin, and Bernhard Schölkopf. A tutorial on ν -support vector machines. *Applied Stochastic Models in Business and Industry*, 21(2):111–136, 2005.
- [19] Bartlett M. Tanaka J. Movellan J. Cockburn, J. and R. Schultz. Smile-maze: A tutoring system in real-time facial expression perception and production in children with Autism Spectrum Disorder. In *In Proceedings from the IEEE International Conference on Automatic Face & Gesture Recognition*, pages 978–986. IEEE, 2008.
- [20] J. F. Cohn and K. L. Schmidt. The timing of facial motion in posed and spontaneous smiles. *J. Wavelets, Multi-resolution & Information Processing*, 2:1–12, 2004.
- [21] Jeffrey F. Cohn, Adena J. Zlochower, James Lien, and Takeo Kanade. Automated face analysis by feature point tracking has high concurrent validity with manual faces coding. *Psychophysiology*, 36:35–43, 1999.

- [22] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In Hans Burkhardt and Bernd Neumann, editors, *Computer Vision ECCV98*, volume 1407 of *Lecture Notes in Computer Science*, pages 484–498. Springer Berlin / Heidelberg, 1998. 10.1007/BFb0054760.
- [23] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59, January 1995.
- [24] T.F. Cootes. Model-based methods in analysis of biomedical images. In R. Baldock and J. Graham, editors, *Image Processing and Analysis*, chapter 7, pages 223–248. Oxford University Press, 2000.
- [25] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):681–685, June 2001.
- [26] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Mach. Learn.*, 20:273–297, September 1995.
- [27] David Cristinacce and Tim Cootes. Automatic feature localisation with constrained local models. *Pattern Recogn.*, 41:3054–3067, October 2008.

- [28] Tomaž Curk, Janez Demšar, Qikai Xu, Gregor Leban, Uroš Petrovič, Ivan Bratko, Gad Shaulsky, and Blaž Zupan. Microarray data mining with visual programming. *Bioinformatics*, 21:396–398, February 2005.
- [29] Fernando De la Torre and Jeffrey F. Cohn. Facial expression analysis. In Th.B. Moeslund, A. Hilton, V. Krger, and L. Sigal, editors, *Guide to Visual Analysis of Humans: Looking at People. Part 3*, pages 377–409. Springer, 2011.
- [30] Jane Reilly Delannoy and John B. McDonald. Estimation of the temporal dynamics of posed and spontaneous facial expression formation using LLE. *International Machine Vision and Image Processing Conference*, 0:139–144, 2009.
- [31] P. Ekman. *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Holt Paperbacks, second edition, 2007.
- [32] P. Ekman. *Telling Lies: Clues to Deceit in the marketplace, politics, and marriage*. W. W. Norton & Company, Inc., reissued edition, 2009.
- [33] 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.

- [34] P. Ekman, W. Friesen, and J. Hager. *Facial Action Coding System - The Manual on CD-ROM*, 2002.
- [35] Paul Ekman. Darwin, deception, and facial expression. *Annals Of The New York Academy Of Sciences*, 1000(1):205–221, 2003.
- [36] Paul Ekman and Richard J Davidson. Voluntary smiling changes regional brain activity. *Psychological Science*, 4(5):342–345, 1993.
- [37] Paul Ekman and W Friesen. Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, 6:238–252, 1982.
- [38] Paul Ekman, Wallace V Friesen, and Maureen O’Sullivan. Smiles when lying. *Journal of Personality and Social Psychology*, 54(3):414–420, 1988.
- [39] Nathan A. Fox and Richard J. Davidson. Patterns of brain electrical activity during facial signs of emotion in 10-month-old infants. *Developmental Psychology*, 24(2):230–236, 1988.
- [40] M. Frank, P. Perona, and Y. Yacoob. Automatic extraction of facial action codes. <http://mplab.ucsd.edu/grants/project1/publications/pdfs/CIAConsultantsReport.pdf>, 2001. Accessed on April 4th, 2012.

- [41] Mark G. Frank, Paul Ekman, and Wallace V. Friesen. Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality & Social Psychology*, 64(1):83 – 93, 1993.
- [42] Isabel Gonzalez, Hichem Sahli, and Werner Verhelst. Automatic recognition of lower facial action units. In *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research*, MB '10, pages 8:1–8:4, New York, NY, USA, 2010. ACM.
- [43] Ralph Gross, Iain Matthews, and Simon Baker. Active appearance models with occlusion. *Image Vision Comput.*, 24:593–604, June 2006.
- [44] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, 2003.
- [45] Takeo Kanade, Yingli Tian, and Jeffrey F. Cohn. Comprehensive database for facial expression analysis. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, FG '00, pages 46–, Washington, DC, USA, 2000. IEEE Computer Society.
- [46] Eva Krumhuber and Arvid Kappas. Moving Smiles: The role of dynamic components for the perception of the genuineness of smiles. *Journal of Nonverbal Behavior*, 29:3–24, 2005. 10.1007/s10919-004-0887-x.

- [47] Eva G. Krumhuber and Antony S.R. Manstead. Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion Washington Dc*, 9(6):807–820, 2009.
- [48] Charles X. Ling, Jin Huang, and Harry Zhang. AUC: A statistically consistent and more discriminating measure than accuracy. In *Proceedings of the 18th international joint conference on Artificial intelligence, IJCAI'03*, pages 519–524, San Francisco, CA, USA, 2003. Morgan Kaufmann Publishers Inc.
- [49] Gwen Littlewort, Marian Stewart Bartlett, Linda P. Salamanca, and Judy Reilly. Automated measurement of children’s facial expressions during problem solving tasks. In *FG*, pages 30–35. IEEE, 2011.
- [50] Gwen Littlewort, Jacob Whitehill, Tingfan Wu, Ian R. Fasel, Mark G. Frank, Javier R. Movellan, and Marian Stewart Bartlett. The Computer Expression Recognition Toolbox (cert). In *FG*, pages 298–305. IEEE, 2011.
- [51] Gwen C. Littlewort, Marian Stewart Bartlett, and Kang Lee. Faces of pain: Automated measurement of spontaneous facial expressions of genuine and posed pain. In *Proceedings of the 9th international conference on Multimodal interfaces, ICMI '07*, pages 15–21, New York, NY, USA, 2007. ACM.

- [52] Gwen C. Littlewort, Marian Stewart Bartlett, and Kang Lee. Automatic coding of facial expressions displayed during posed and genuine pain. *Image Vision Comput.*, 27:1797–1803, November 2009.
- [53] Gwen Littlewort-Ford, Marian Stewart Bartlett, and Javier R. Movellan. Are your eyes smiling? Detecting genuine smiles with support vector machines and Gabor wavelets. In *Proceedings of the 8th Joint Symposium on Neural Computation*, 2001.
- [54] P. Lucey, J.F. Cohn, T. Kanade, J.M. Saragih, Z. Ambadar, and I. Matthews. The Extended Cohn-Kanade Dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *CVPR4HB10*, pages 94–101, 2010.
- [55] Simon Lucey, Ahmed Bilal Ashraf, and Jeffrey Cohn. Investigating spontaneous facial action recognition through AAM representations of the face. In K. Kurihara, editor, *Face Recognition Book*. Pro Literatur Verlag, Mammendorf, Germany, April 2007.
- [56] Iain Matthews and Simon Baker. Active appearance models revisited. *Int. J. Comput. Vision*, 60:135–164, November 2004.
- [57] M. Pantic, M. F. Valstar, R. Rademaker, and L. Maat. Web-based database for facial expression analysis. In *Proceedings of IEEE Int’l*

- Conf. Multimedia and Expo (ICME'05)*, pages 317–321, Amsterdam, The Netherlands, July 2005.
- [58] E.K. Patterson and A. Gaweda. Toward using dynamics of facial expressions and gestures for person identification. In *Proceedings of the Fifth IASTED International Conference*, volume 711, pages 026–58, 2010.
- [59] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572, 1901.
- [60] Matthew Ratliff. Active appearance models for affect recognition using facial expressions. Master’s thesis, University of North Carolina Wilmington, 601 South College Road, Wilmington, North Carolina, USA, 2010.
- [61] Irina Rish. *An empirical study of the naive Bayes classifier*, pages 41–46. 2001.
- [62] E L Rosenberg, P Ekman, and J A Blumenthal. Facial expression and the affective component of cynical hostility in male coronary heart disease patients. *Health Psychol*, 17(4):376–80, 1998.
- [63] Jason M Saragih, Simon Lucey, and Jeffrey F Cohn. Face alignment through subspace constrained mean-shifts. *Computer Vision 2009 IEEE 12th International Conference on*, (CIm):1034–1041, 2009.

- [64] Arman Savran, Neşe Alyüz, Hamdi Dibeklioglu, Oya Çeliktutan, Berk Gökberk, Bülent Sankur, and Lale Akarun. Biometrics and identity management. chapter Bosphorus Database for 3D Face Analysis, pages 47–56. Springer-Verlag, Berlin, Heidelberg, 2008.
- [65] Arman Savran, Bülent Sankur, and Taha M. Bilge. Estimation of facial action intensities on 2D and 3D data. In *Proceedings of the 19th European Signal Processing Conference (EUSIPCO 2011)*, Barcelona, Spain, August 2011.
- [66] K.L. Schmidt, A. Gentry, J.K. Monin, and K.L. Courtney. Demonstration of facial communication of emotion through telehospice videophone contact. *Telemed J E Health*, 17(5):399–401, 2011.
- [67] Bernhard Schölkopf, Alex J. Smola, Robert C. Williamson, and Peter L. Bartlett. New support vector algorithms. *Neural Comput.*, 12(5):1207–1245, May 2000.
- [68] N. Sebe, M.S. Lew, Y. Sun, I. Cohen, T. Gevers, and T.S. Huang. Authentic facial expression analysis. *Image and Vision Computing*, 25(12):1856–1863, 2007.
- [69] Lindsay I. Smith. A tutorial on principal component analysis. http://www.cs.otago.ac.nz/cosc453/student_tutorials/

- principal_components.pdf, February 2002. Accessed on April 4th, 2012.
- [70] Veikko Surakka and Jari K Hietanen. Facial and emotional reactions to Duchenne and non-Duchenne smiles. *International Journal of Psychophysiology*, 29(1):23–33, 1998.
- [71] Ying-li Tian, Takeo Kanade, and Jeffrey F. Cohn. Multimodal interface for human-machine communication. chapter Recognizing action units for facial expression analysis, pages 32–66. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2002.
- [72] Yan Tong, Jixu Chen, and Qiang Ji. Modeling and exploiting the spatio-temporal facial action dependencies for robust spontaneous facial expression recognition. *Computer Vision and Pattern Recognition Workshop*, 0:34–41, 2009.
- [73] Yan Tong, Jixu Chen, and Qiang Ji. A unified probabilistic framework for spontaneous facial action modeling and understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:258–273, 2010.
- [74] M. F. Valstar and M. Pantic. Induced disgust, happiness and surprise: An addition to the mmi facial expression database. In *Proceedings of Int’l Conf. Language Resources and Evaluation, Workshop on EMOTION*, pages 65–70, Malta, May 2010.

- [75] Michel F. Valstar, Hatice Gunes, and Maja Pantic. How to distinguish posed from spontaneous smiles using geometric features. In *Proceedings of the 9th international conference on Multimodal interfaces, ICMI '07*, pages 38–45, New York, NY, USA, 2007. ACM.
- [76] Michel F. Valstar, Maja Pantic, Zara Ambadar, and Jeffrey F. Cohn. Spontaneous vs. posed facial behavior: Automatic analysis of brow actions. In *Proceedings of the 8th international conference on Multimodal interfaces, ICMI '06*, pages 162–170, New York, NY, USA, 2006. ACM.
- [77] Vladimir N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.
- [78] Esra Vural, Marian Bartlett, Gwen Littlewort, Mujdat Cetin, Aytul Ercil, and Javier Movellan. Discrimination of moderate and acute drowsiness based on spontaneous facial expressions. *Pattern Recognition, International Conference on*, 0:3874–3877, 2010.
- [79] Kilian Q. Weinberger, John Blitzer, and Lawrence K. Saul. Distance metric learning for large margin nearest neighbor classification. In *In NIPS*. MIT Press, 2006.
- [80] Jacob Whitehill, Marian Bartlett, and Javier Movellan. Measuring the perceived difficulty of a lecture using automatic facial expression recognition. In *Proceedings of the 9th international conference on Intelli-*

gent Tutoring Systems, ITS '08, pages 668–670, Berlin, Heidelberg, 2008.
Springer-Verlag.

- [81] Harry Zhang. The optimality of naive Bayes. In Valerie Barr and Zdravko Markov, editors, *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference*, volume 1, page 3. AAAI Press, 2004.

A Confusion Matrices

		Predicted Class		Total
		Non-Duchenne	Duchenne	
Actual Class	Non-Duchenne	109	14	123
	Duchenne	31	85	116
Total		140	99	239

Table 3: *Ideal-Case* Approach - Confusion Matrix - k -Nearest Neighbor

		Predicted Class		Total
		Non-Duchenne	Duchenne	
Actual Class	Non-Duchenne	96	27	123
	Duchenne	20	96	116
Total		116	123	239

Table 4: *Ideal-Case* Approach - Confusion Matrix - Naïve Bayes

		Predicted Class		Total
		Non-Duchenne	Duchenne	
Actual Class	Non-Duchenne	115	8	123
	Duchenne	5	111	116
Total		140	99	239

Table 5: *Ideal-Case* Approach - Confusion Matrix - Support Vector Machine

		Predicted Class		Total
		Non-Duchenne	Duchenne	
Actual Class	Non-Duchenne	108	15	123
	Duchenne	23	93	116
Total		131	108	239

Table 6: *Semi-Generalized* Approach - Confusion Matrix - k -Nearest Neighbor

		Predicted Class		Total
		Non-Duchenne	Duchenne	
Actual Class	Non-Duchenne	93	30	123
	Duchenne	13	103	116
Total		106	133	239

Table 7: *Semi-Generalized* Approach - Confusion Matrix - Naïve Bayes

		Predicted Class		Total
		Non-Duchenne	Duchenne	
Actual Class	Non-Duchenne	113	10	123
	Duchenne	3	113	116
Total		116	123	239

Table 8: *Semi-Generalized* Approach - Confusion Matrix - Support Vector Machine

		Predicted Class		Total
		Non-Duchenne	Duchenne	
Actual Class	Non-Duchenne	57	6	63
	Duchenne	13	43	56
Total		70	49	119

Table 9: *Fully-Generalized* Approach - Confusion Matrix - k -Nearest Neighbor

		Predicted Class		Total
		Non-Duchenne	Duchenne	
Actual Class	Non-Duchenne	44	19	63
	Duchenne	12	44	56
Total		56	63	119

Table 10: *Fully-Generalized* Approach - Confusion Matrix - Naïve Bayes

		Predicted Class		Total
		Non-Duchenne	Duchenne	
Actual Class	Non-Duchenne	56	7	63
	Duchenne	2	54	56
Total		58	61	119

Table 11: *Fully-Generalized* Approach - Confusion Matrix - Support Vector Machine