

2010

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

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

INDIVIDUAL IDENTIFICATION USING DYNAMIC FACIAL EXPRESSIONS
WITH ACTIVE APPEARANCE-BASED HIDDEN MARKOV MODELS

Adam Matthew Gaweda

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

2010

Approved by

Advisory Committee

Karl Ricanek

Ulku Yaylacicegi

Eric Patterson

Chair

Accepted by

Dean, Graduate School

TABLE OF CONTENTS

1	ABSTRACT	iii
2	ACKNOWLEDGEMENTS	iv
3	INTRODUCTION	1
4	BACKGROUND MATERIAL	4
	4.1 PSYCHOLOGY AND ANTHROPOLOGY	4
	4.2 EXPRESSION RECOGNITION	5
	4.3 MOTIVATION	6
	4.4 OTHER AREAS OF DYNAMIC RECOGNITION-RELATED RESEARCH	7
5	FUNDAMENTAL TECHNOLOGIES	9
	5.1 PROTOTYPE EXPERIMENT	9
	5.2 PRINCIPAL COMPONENT ANALYSIS	9
	5.3 ACTIVE SHAPE MODELS	10
	5.4 ACTIVE APPEARANCE MODELS	11
	5.5 HIDDEN MARKOV MODELS	14
6	EXPERIMENT	17
	6.1 PHYSICAL DESIGN	17
	6.2 DATA ACQUISITION	17
	6.2.1 EXPRESSIONS USED	19
	6.4 DATA EXTRACTION	19
	6.4.1 STACKED ACTIVE SHAPE MODELS	20
	6.4.2 DETECTION OF THE FACE - VIOLA-JONES AND ROWLEY	21
	6.4.3 DETECTION OF FACIAL FEATURES	22
	6.4.4 ACTIVE APPEARANCE MODELS	24
	6.5 DATA ANALYSIS	24
	6.5.1 HIDDEN MARKOV MODELS	24
7	FINDINGS	26
	7.1 EXPERIMENTS	26
	7.1.1 STATIC COEFFICIENTS	26
	7.1.2 STATIC-DELTA COEFFICIENTS	27
	7.1.3 DELTA-ONLY COEFFICIENTS	27
	7.1.4 DELTA-ACCELERATION COEFFICIENTS	29
	7.1.5 COMBINED DELTA COEFFICIENTS	33
	7.1.6 SINGLE EXPRESSION DELTA COEFFICIENTS VER- SUS COMBINED DELTA COEFFICIENTS	37
	7.2 OBSERVATIONS	38
8	FUTURE WORK	39
9	CONCLUSION	41
	REFERENCES	43

ABSTRACT

Individual Identification using Dynamic Facial Expressions with Hidden Active Appearance Markov Models. Adam Matthew Gaweda, 2010. Thesis Paper, University of North Carolina Wilmington.

Determining identity is becoming an increasingly important and heavily researched area of computational intelligence. Typically measurable biological characteristics, or biometrics, are used to quantify the physical features of an individual in order to use them as a means of identification. Common biometrics, such as fingerprints, the iris, and one's voice, assist in the determining process. There have been psychological studies recently that indicate a new biometric, body language with a focus on the expressions made from the face, could be used. In this work, the hypothesis is that facial dynamics of an individual face could be used as an effective biometric for person identification. The method described here applies Stacked Active Shape Models for automated face detection and labeling, Active Appearance Models for feature extraction, and Hidden Markov Models for data analysis. Individual models are constructed for each person in this scenario and used to test identification with new video of facial expressions of the same individuals. Results confirm the hypothesis and demonstrate the efficacy of the potential approach.

ACKNOWLEDGEMENTS

First and foremost I would like to acknowledge the support of my family. I was introduced to the family computer in 1993 and my interest in technology has gone since. My father, working two jobs to support our family, taught me the importance of hard work on any personal goal. The family business, an irrigation company, is where I learned to work and save for something I wanted. My mother has supported every choice that I have made throughout my life, through some of the calmest times to the harshest. My mother encouraged technology in the home, and encouraged learning computers to join the emerging industry. To my three sisters, Ginny, Stacey, and Sarah, I was raised to be both the older and younger brother to them. Their friends were my friends and have always been there when I needed them.

Secondly, I would like to thank the Kure Beach Aikido organization and the Aikido World Alliance as a whole for all the support they have given me during the near-four years that I have been practicing the art. It was through Aikido that I found an interest in human interactions, and it is a privilege to hear the insight of personal interaction from the psychologists, doctors, and scholars that I have met throughout my training.

Finally, I would like to thank the UNCW Face Aging Group for the opportunity to work on such a specialized project during my time as a graduate student. Without their help, along with the individuals that allowed their faces to be recorded, I would not have had the privilege to develop and showcase my abilities as a researcher. The technologies of the future are truly here today, and it was great deal of enjoyment and fascination actively participating in the progression of the human computer interface.

LIST OF FIGURES

1	Timeline of Events	3
2	Example of a smile. Using FACS scoring, this is classified as 6D+7C+12E+25B	6
3	The shape model used in prototype experiment.	9
4	The texture map extracted from prototype experiment.	9
5	Example of a labeled subject’s face using 23 landmark points.	10
6	Example of Active Shape Models. This shape is trained across the labeled images from the database.	11
7	Active Appearance Models are represented by eigenvectors that are the distance between the mean image.	12
8	First 5 modes of the Active Appearance Models.	13
9	Physical Set-Up of Experiment	17
10	Each frame was mapped with 161 points to categorize the face and facial features.	18
11	The expressions shown to subjects in order for them to mimic. From left to right: mad, happy, sticking the tongue out, sad, and surprise.	19
12	Example of a test subject’s expressions	20
13	Example of Stasm’s Initial Detection and Convergence	21
14	Viola-Jones employs edge detection from Haar-like feature vectors like above to detect the face.	22
15	Example of the ‘whiskers’ used by Stasm. These whiskers are orthogonal to the shape model.	23
16	The ‘whiskers’ intersect with the shape model. Each intersection is centered at a small positive or negative displacement along the whisker, typically offset by 3 pixels.	23
17	A delta regression formula is used to calculate the difference between to frames	25
18	Results of Static-Only Coefficients Testing based and State and Expression	28
19	Results of Delta-Only ‘Mad’ Tests aligned by States	29
20	Results of Delta-Only ‘Happy’ Tests aligned by States	30

21	Results of Delta-Only ‘Sticking-Out Tongue’ Tests aligned by States	30
22	Results of Delta-Only ‘Sad’ Tests aligned by States	31
23	Results of Delta-Only ‘Surprise’ Tests aligned by States	31
24	Results of Delta-Only ‘Business’ Tests aligned by States	32
25	Results of Delta-Only ‘Pleasure’ Tests aligned by States	32
26	Results of Delta-Acceleration ‘Mad’ Tests aligned by States	33
27	Results of Delta-Acceleration ‘Happy’ Tests aligned by States	34
28	Results of Delta-Acceleration ‘Sticking-Out Tongue’ Tests aligned by States	34
29	Results of Delta-Acceleration ‘Sad’ Tests aligned by States	35
30	Results of Delta-Acceleration ‘Surprise’ Tests aligned by States	35
31	Results of Delta-Acceleration ‘Business’ Tests aligned by States	36
32	Results of Delta-Acceleration ‘Pleasure’ Tests aligned by States	36
33	Results of Combined Delta Coefficients sorted by State	37
34	Results of Single Expression Delta Coefficients Combined Delta Coefficients sorted by State	38

INTRODUCTION

Person identification is an important area of security. Under security, access control handles authentication, authorization, and audit. In order to determine whether an individual has the desired authorization, access controls are in place to limit the availability of resources or areas. Access controls can be as simple as a fence around the perimeter of an area to a lock on a door to a user's login password. One form of access control is to use physical features in order to positively identify someone. In early systems, a person's face was used for identification. As technology increased, other physical features, like thumbprints or the iris, were used in conjunction with the face to increase accuracy. This method of using physical features for the purpose of identification is known as biometrics.

All biometrics are governed by three important key factors: universality, distinction, and repeatability. To be universal, the biometric needs to be the same for almost all cultures, such as an iris or fingerprint. Considerably the most common physical feature used in human biometrics is the face. One of the most defining features of the human body, the face is generally the first physical feature other individuals will recognize. In all cultures the face generally contains two eyes, resting on the elongated nose, which sits atop the mouth. Because almost no two faces are alike, exception in special cases, the face is considered distinct to each person. Finally, the face is considered repeatable because on day-to-day basis, it does not generally change except from internal mechanisms such as expressions or aging; or external mechanisms such as cosmetic surgery or injury.

Biometrics are commonly used in many business aspects as a security protocol for access control. Facial recognition software allows specific users to be given access. In some cases, the face is not a sufficient control mechanism. Vocal patterns are used as well to identify someone for access. Both instances can still have a degree of error by uncontrollable elements (occlusions in the face, and background noise in voice). In military security, sometimes an uninvasive method is needed to identify individuals from a distance. In order to handle

this, various biometrics are investigated to overcome these obstacles, such as gait analysis. Gait analysis, using one's body movements as a biometric, can be used in these cases to ensure that a person's face matches their body movements. With multiple positive identifications, it becomes harder to create a false-positive, identifying someone who they are not.

Many biometrics, however, do not focus solely on the physical features of the body to identify individuals. Use of body language has long since been used as a way to visually lip read or identify individuals whose features may otherwise be occluded[20]. Body language also is known to fulfill the three criteria of a biometric. Gestures can be considered universal due to the similarities in the facial muscles used for expressions. These same facial muscles help shape the face and are used in ways that are considered unique to each person. Finally, gestures can be repeated multiple times over, allowing one's body language to serve as a biometric.

One of the possibilities of improving biometrics is the use of dynamic information. By merging the biometric capabilities of the face with the unique expressions of the face, a new dynamic biometric can be used in an attempt to increase accuracy of identification. Including any dynamics of the face at all in automated methods is a relatively new area of study, largely due to the capability of systems to locate, extract, and process the required data. Once expressions and gestures of the face are located, captured, and classified, individuals may be more robustly identified in real-world scenarios as well as have classifications made about indicators to behavior and intent. This extends recent work in psychology and anthropology to bring new capabilities to automated pattern recognition systems.

The term "dynamics", in regards to biometrics, focuses on the changes in motion over sequential time of the body. In the process of a person raising one's hand in the air, the "dynamic" would be considered the body motion from the initiation of the arm raise to the completed of the act, that is, either returning the arm back to a neutral state or reaching the peak of the raise.

This paper's hypothesis is that facial dynamics of an individual face could be used as an effective biometric for person identification. This

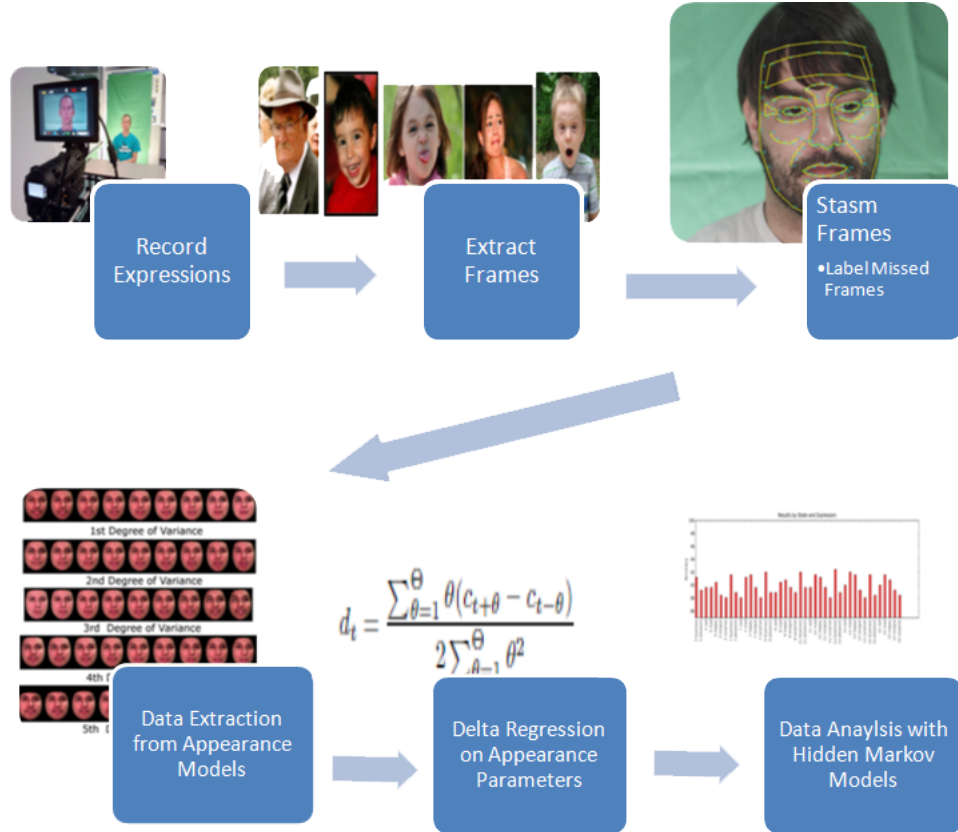


Figure 1: Timeline of Events

hypothesis will be tested by constructing and testing a prototype facial dynamic biometric system. We present a brief survey of related work and propose an initial system for using facial gestures, using various expressions, as a biometric for human identification. The methods presented apply a technique known as Active Appearance Models (AAM), AAM features to Hidden Markov Model (HMM) classification. The AAM landmarks in this experiment focus on two dimensional representations of the face captured from a single camera facing directly at the subject. The approach we propose for identification by face movement is described, and experiments are presented to demonstrate feasibility with noteworthy initial results. Results from the experiment are detailed, and discussion is given toward current and future work.

BACKGROUND MATERIAL

We present the goal of creating automated biometric systems that are capable of dynamically identifying and classifying facial expressions and gestures for both person identification and also for classification of expressions. In considering this, there are several areas of related research that can provide useful methods and information. In this section, we survey some of the important related work in psychology, anthropology, and computing in order to present a brief summary of the current state of the field.

PSYCHOLOGY AND ANTHROPOLOGY

Since the times of Descartes, there has been investigation over the universality of human emotions. Descartes is even considered one of the forefathers of emotion investigation simply because ‘no one had written on these matters before’ [11]. Following Descartes’ classification of the characteristics of emotions, philosophers Spinoza and Hobbes took a radical approach to considering emotions as the result of man’s inadequate understanding of life [51, 25]. It was not until naturalist Charles Darwin hypothesized that the expressions commonly made when an individual is under the influence of a given emotion were considered to be universal [9].

There are 43 muscles in the face, most of which are controlled by the seventh cranial nerve (also known as the facial nerve). This nerve exits the cerebral cortex and emerges from your skull just in front of your ears. It then splits into five primary branches: temporal, zygomatic, buccal, mandibular and cervica[56]. It is with the flexibility of these muscles that we can create facial expressions. Studies conducted by Duchenne catalogued many expressions, generated by inducing electrical shocks through ‘electropuncture’ into the facial muscles of subjects, followed by photographing the resulted expression. The term, Duchenne smile, is attributed to his research, discovering true emotional happiness contained stimulated muscles in not only the mouth region of the face, but also the ocular region. In *The Expression of the Emotions in Man and Animals*, Darwin’s drawings are all inspired by the photographs taken during Duchenne’s experiments[49, 12].

In psychological studies, research has shown the humans have the ability to positively identify people based on their gait, or body mechanics. O’Toole and Roark discussed that facial dynamics could support person identification when handling various viewing conditions[39, 40]. Hill and Johnston furthered this investigation by reporting that humans were able to identify individuals after looking a 3-dimensional character that held no physical resemblance to the sources[24].

EXPRESSION RECOGNITION

Research has begun to explore recognition of emotions based on the exploits of Duchenne’s studies. Emotional classification has expanded from six basic emotions (Happy, Sad, Angry, Surprised, Fear, and Disgust) [18] to robust, complex combinations of emotions (Amusement, Shame, Guilt, Satisfaction, etc.) [7]. Robert Plutchik designed a model of the ranging extremities of emotions. By combining two adjacent emotions on the model, Plutchik showed that a third, complex emotion would be the result [44]. Paul Ekman first categorized the six emotions previously mentioned, then added thirteen additional emotions[17, 7].

Ekman also introduced the Facial Action Coding System (FACS) as a way to help study and classify facial movements in expressions. FACS uses Action Units (AU) to ‘score the muscular activity that produces momentary changes in facial appearance’ [19]. This score specifies the muscle group that is tightening along with the ranking scale from A to E (low to extreme, respectfully) to describe the level of intensity of the muscles tightened.

In more recent years, studies have taken facial mapping systems and used them for purposes of identification. In today’s society, there is a number of different facial recognition software systems used for research and for security. Due to these techniques using single still images to identify, there are serious limitations to the system; lighting concerns can hide areas of the face, occlusion of the individual’s face (hats, glasses, facial hair, etc.) can hide key identifiable elements, and various angles of the face can outright cause many face detecting systems to stop working[20].

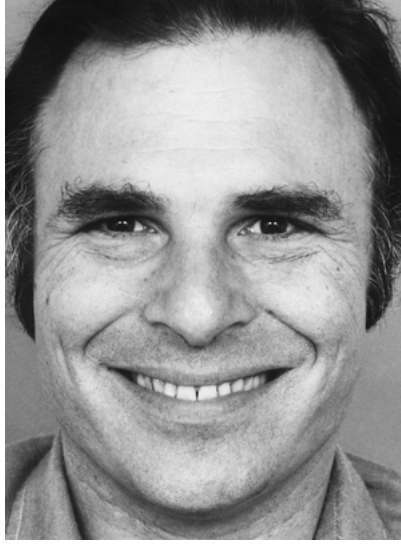


Figure 2: Example of a smile. Using FACS scoring, this is classified as 6D+7C+12E+25B

For this paper, many of the emotions initially described by Paul Ekman were used to classify the expressions captured from individuals. The use of FACS, however, was not used due to the computational complexity of the coding system as well as the possibility of it limiting fine-grain detail available in tracking a large number of specific face landmarks.

MOTIVATION

While facial dynamics may seem to be a novel biometric, it is being investigated under serious cases of occlusion. Many modern systems require the face or at least the iris to be visible at the time of data collection. If the data is unable to extract the face properly, then many recognition systems will fail. Another case of being unable to tell two parties apart can be from twins or body doubles. Since biometric systems are meant to verify specific individuals, the ultimate test for the system is when another person looks identical or near-identical to the authorized person. There is now a thorough investigation underway to what physical aspects are still unique among look-a-likes[28, 30, 52].

Another area of investigation is observing the ‘microexpressions’ of individuals. These subtle movements can be classified for any number of reasons from emotion detection to deception detection. Paul Ekman

has investigated the nuances of the face in relation to lying multiple times[16]. If the recognition system is able to use the subtle gestures of the face to identify individuals, the next area of research would be to develop a system that ‘picks up’ the microexpressions of individuals. Since deception is a subjective act that can vary from person to person, being able to identify commonly seen expressions that are indictative of deception can be used to identify moments in a person’s interaction in which deception, or least defensive conversation, was displayed.

The use of Hidden Markov Models to verify the movements of the human face for recognition has been suggested, and only recently investigated[29, 33, 53]. HMMs allow for sequential data, such as changes from the face over time, to represent the biometric information needed for verification. Unfortunately, with higher quality footage, the number of images used increases. Depending on the number of landmark points used in the extraction process, the labeling process can become exponentially impossible. Stasm allows most of the images to be automatically labeled. Active Appearance Models cause representation of each face’s information to be compared to the collective average.

OTHER AREAS OF DYNAMIC RECOGNITION-RELATED RESEARCH

Using facial dynamics for identification is actually closer towards gait recognition than face recognition. Facial dynamics focuses on the movement of the muscles on the face, rather than the texture of the skin. Traditional gait recognition systems analyze the movement of the body in space (kinematics) and the forces involved in producing these movements (kinetics)[23]. Medically, gait is typically used determine posture or where severe impact on joints is occurring[32]. O’Toole’s analysis uses a computer-generated avatar to show that humans can use the gait obtained from the face in order to postively identify people[40]. There has also be investigation of using gait analysis centered around facial dynamics in order to lip read[37, 38].

Audio-visual speech processing has also grown to into a similar field as facial dynamics. In order to handle issues in audio-only recognition systems, multimodel speech recognition systems include video

of speakers in order to increase accuracy of recognition. Techniques for speaker-independent speech reading are also needed. These may include some form of speaker adaptation, model adaptation, image warping, use of feature dictionaries, or some of the many other methods employed for audio speaker adaptation[42, 35, 22]. While visual speaker identification and facial dynamic biometrics are similar, facial dynamics seeks to capture all the differences of the face, instead of focusing on just the mouth.

Facial dynamics do not focus as strongly on the kinematics as the kinetics since the video camera is focused on the face. Traditional face detection systems locate the face within the camera's frame, and with Stasm, label the area to analyze and extract from.

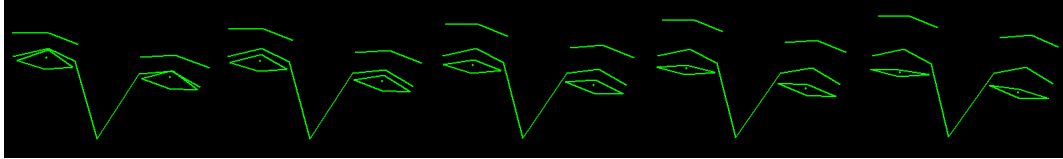


Figure 3: The shape model used in prototype experiment.

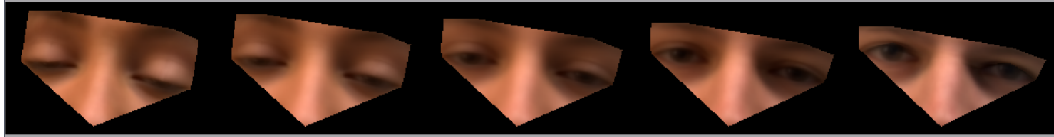


Figure 4: The texture map extracted from prototype experiment.

FUNDAMENTAL TECHNOLOGIES

PROTOTYPE EXPERIMENT

In order to show this technique, prototype tests were first conducted on a smaller scale[41]. Subjects were interviewed about various memories in their lives, and then were asked to deceive the interviewer with a false story. The interviewer was never told which stories were true and which were false. Original motivation was to examine the microexpressions, or ‘leakages’ discussed by Paul Ekman in an attempt to identify degree of deception in the face [15]. From this dataset, we extracted blinks from 12 subjects in order to identify them from their expression dynamics. Since the focus of the experiment was in the periocular region, only 23 landmark points were labeled across the face (See Fig 5).

The parameters were extracted with a delta regression formula (discussed in section). The hidden Markov models were then trained and tested in forced alignment verification process (also discussed in section).

PRINCIPAL COMPONENT ANALYSIS

Before explaining Active Shape and Appearance Models, Principal Component Analysis (PCA) will be discussed, as it is important to Active Shape Models. PCA is a common technique for finding patterns in



Figure 5: Example of a labeled subject's face using 23 landmark points.

data of high dimension. Developed by Karl Pearson, Principal Component Analysis creates predictive models in order to derive the variability of the data given[43, 50]. Depending on the area of research, PCA also goes by the name KarhunenLove transform (KLT), the Hotelling transform or proper orthogonal decomposition (POD)[34, 26]. The stronger variables (ones that hold a higher degree of variance) are organized as the first axis in the new PCA-transformed data space. Subsequent variables with high levels of variance are considered second, third, and so forth. This is only a transformation at this point. It becomes a reduction if you 'throw away' axes, however it allows for efficient representation. Active Shape and Appearance Models utilize PCA order in to find the best representation of shape and textures for the face.

ACTIVE SHAPE MODELS

Active Shape Models (ASM) are statistical models based on the shapes of objects which iteratively manipulate themselves to fit an object in a new image [6]. In order for the statistical models to properly



Figure 6: Example of Active Shape Models. This shape is trained across the labeled images from the database.

adjust to a new image, the shapes are trained by a Statistical Shape Model [2]. A statistical shape model is made through the creation of a set of n points. This set of points is used to create the average location of a point with relation to the flexibility, or variance, of the point in relation to other neighboring points. An example would be the variance in human head sizes may expand or collapse to left and right-most points, but they location of the bottom-most points would stay roughly the same.

ASM relies heavily on strong edge differences to properly locate an object within an image. With an iterative approach, ASM follows a simplistic 3 step approach toward finding the appropriate shape of an object [6]:

- For each point X_i , find the best local match for X_i .
- Update the shape parameters to best fit the model to the found points.
- Repeat these steps until the object and the image reach convergence based on a given cost measure.

ACTIVE APPEARANCE MODELS

As an extension of Active Shape Models, Active Appearance Modeling was originated by Cootes, Edwards, and Taylor as an attempt to classify and interpret images at a high level [5]. The model is built using principal components analysis on both shape and texture vectors of an image, and then again on a combined feature vector to create

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}$$

Figure 7: Active Appearance Models are represented by eigenvectors that are the distance between the mean image.

an appearance model which can then be used to analyze like images as well as categorize information about that image. It’s advantage over older techniques such as ‘eigenfaces’ is component level registration and representation. AAMs have been reported successful in use for a wide variety of techniques such as pose-correction, face tracking, face recognition as well as higher-level tasks such as age synthesis, age/gender/ethnicity classification, beauty modification, and expression recognition [47, 31]. Compared to Active Shape Models, AAMs show a greater level of detail in the representation of the face [4]. While shape models would allow for the tracking of the facial outline based on the expression given, appearance models capture the textures that a shape model would be unable to represent, such as furrowing of the brow or dimples in the cheeks.

In order for Active Appearance Models to accurately parameterize the landmarks of each face, Delaunay Triangulation is performed on the set of landmark points. Delaunay Triangulation [5], originated by Boris Delaunay in 1934, attempts to maximize the minimum angle of all angles of the triangles in the triangulation. Based on the Delaunay condition, the algorithm seeks to create circumcircles through all combinations of three points (that is, create a unique circle that passes through each of the triangle’s vertices). Once all triangles are generated, minimization can be computed. Extreme isosceles triangles are generally avoided as well. Triangulation is used in order to warp the triangular sections to the average shape using bilinear transformation. Once the sections are warped to the average shape, the image’s texture can be averaged[10].

The Active Appearance Model generates an initial average active shape model based on the location of all the training set’s landmark points. Delaunay Triangulation is once again applied to this average active shape model to create a reference to the general location of any

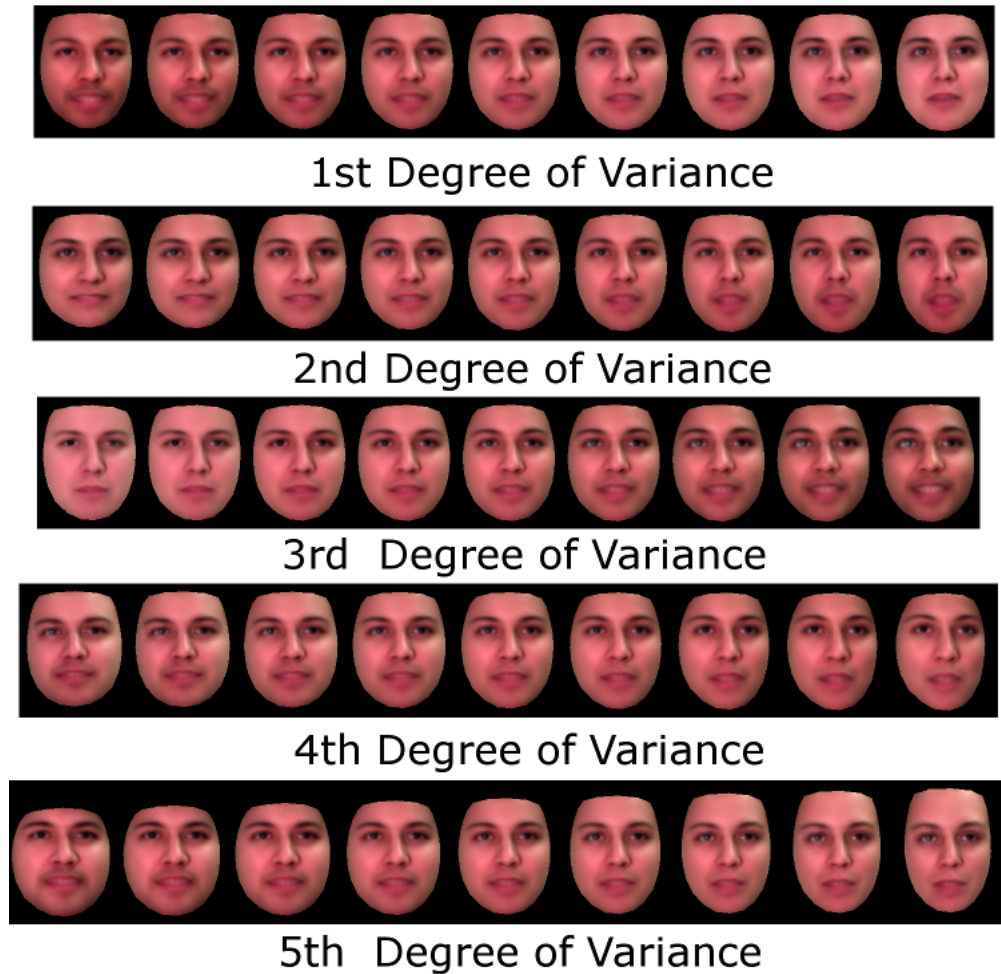


Figure 8: First 5 modes of the Active Appearance Models.

given triangle. The original images' triangular textures may then be warped to fit the shape model. Vectors for the shape and texture are combined and then principal component analysis is applied again.

Finally, the Active Appearance Models generate eigenvectors to represent individual samples in relation to the combined texture and shape model. These vectors act as the variance of the image, thus, adding the vectors to the combined model would recreate the image.

Since the conception of AAM, various fields (such as computer vision, computer-aided aging, medical searches, etc.) have adopted the technique, demonstrating the capability of making highly representative models that can deal with subtle changes in the face and those dynamics as well. Also, while this experiment does not take advantage

of this element, 3-dimensional models could be created to create an even more robust system [1].

HIDDEN MARKOV MODELS

Hidden Markov models are one of the most successful dynamic classification techniques and have been employed for many years with success for speech recognition, speaker identification, and more recently, visual lip-reading as well as other areas of pattern recognition [57]. HMMs are probabilistic state machines built on the assumption that the current state depends only on the previous state and not on the chain of events up to that time [45]. This model can be applied to the dynamics of expressions because the ‘sample’ of a facial movement in a given frame can give directional information (such as the eyes narrowing or widening) based on information from just the previous frame. Each state of an HMM has a probability distribution function based on the data that it represents, and states have probabilities of transitioning to any other states (depending on the model). Based on these aspects, a recognition system can be created that is similar to that of an audio-recognition system.

Each individual has a model that is trained on a series of expressions described earlier. When an expression is tested, each model is traversed, and the model yielding the highest output probability is chosen to identify the individual. Upon collecting the data from appearance modeling, the parameters were formulated into a Hidden Markov Model (HMM). Markov Models have been used in the past as a way to distinguish a potential future state, given the data’s present state [13]. The state is visible and therefore the state transition probabilities are the only parameters. In a HMM, the state is not always transparent, though output dependence of the state is visible. Each state has a probability distribution over all of the possible outputs. Hidden Markov has been employed to aid in the analysis of pattern recognition in voice and gestures, among other fields of study. Therefore, by knowing the sequence of observations, the state should be available as well.

Upon initializing data, Hidden Markov Models divide the training observation vectors equally among the model states and calculates

the mean and variance of each vector. It reassigns the observation vectors to states and recalculates the mean and variance using Baum-Welsh re-estimation. This process is repeated until the estimates find equilibrium and do not change [57].

In order to determine the HMM parameters for each model, a rough guess is made. This rough guess is processed by the Baum-Welsh re-estimation formula in order to narrow down the correct parameters. The steps in this algorithm may be summarised as follows:

- For every parameter vector/matrix requiring re-estimation, allocate storage for the numerator and denominator summations of the form. These storage locations are referred to as accumulators.
- Calculate the forward and backward probabilities for all states j and times t .
- For each state j and time t , use the probability $p(t)$ and the current observation vector $o(t)$ to update the accumulators for that state.
- Use the final accumulator values to calculate new parameter values.
- If the probability for this iteration is not higher than the value at the previous iteration then stop, otherwise repeat the above steps using the new re-estimated parameter values.

The above step assume a single observation sequence; however, for better results, multiple observations should be used. During the experiments conducted, each model was trained on three observations, then tested on a new observation.

Finally, the HMM finds the maximum likelihood state sequence using the Viterbi algorithm. The Baum-Welsh re-estimation formula also creates a total likelihood probability. This probability is used for recognition when processed with the Viterbi's maximum likelihood value. This algorithm finds the best path through a matrix representing the likelihood of transitioning between states. The total probability of any path is computed by summing the transitional probabilities with

the output probabilities along the path. The path with the highest value, or maximum likelihood, is considered identified model.



Figure 9: Physical Set-Up of Experiment

EXPERIMENT

PHYSICAL DESIGN

To capture the dynamics of the face requires fine analysis of the minute changes in the face as the expression occurs. In traditional cameras, the outline of facial features can become blurred and with poor resolution results in confusion between where one feature ends and another begins. In conjunction with the Renaissance Computing Institute (RENCI), individuals' expressions were recorded using a RED ONE camera[46].

DATA ACQUISITION

Each individual was recorded at 100 frames per second (fps) at 2K resolution (2048 x 1152) with a 50mm lens. Subjects sat 7 to 9 feet away from the camera lens across a backdrop. To capture specific dynamics for the purpose of testing, individuals were shown 5 separate images of people giving distinct expressions and then 2 words ('Business' and 'Pleasure'). They were instructed to mimic the expression

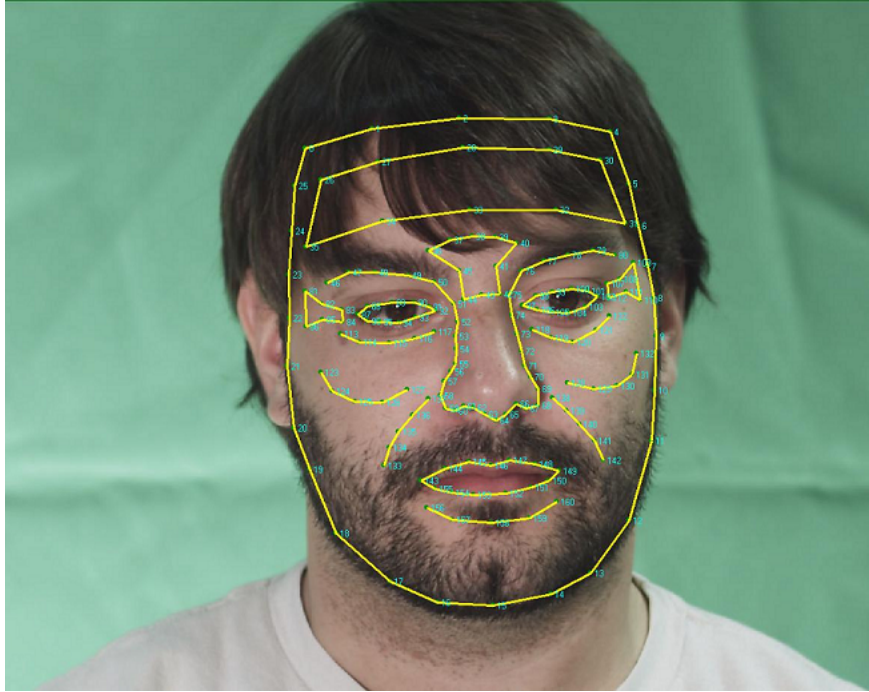


Figure 10: Each frame was mapped with 161 points to categorize the face and facial features.

that was shown to them to the best of their ability (for example, when the subject was shown an image of a happy children, they were asked to mimic the smile the child was showing). When presented with the words, the person was asked to read the word aloud.

The videos were rendered on a Mac OS 10.6.3 workstation using *RED-Alert!* software used to process the RED ONE’s proprietary video format (R3D)[46]. The images were converted from a 2K resolution TIFF image to a 1024 x 576 resolution JPEG image. This conversion was only done to speed up the image processing operations of the experiment and should have no other effect on the results of the experiment. While the whole of the experiment was conducted on the workstation, a parallel Windows OS virtual machine was installed on the workstation and also used in the experiment due to technical requirements for Cootes’ AAM software package, which was used to extract the appearance parameters.

The database created for this experiment resulted in 63 individual recordings. Of the subjects, 33 were male and 32 were female. While there was diversity of ethnicity in the database, over 75% of the

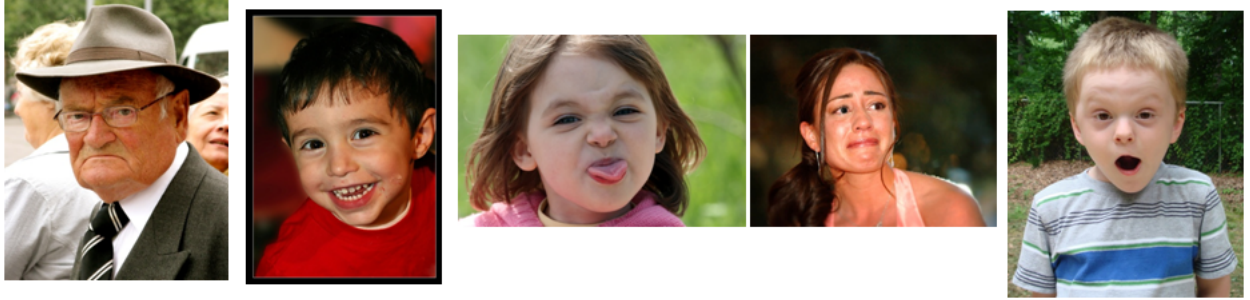


Figure 11: The expressions shown to subjects in order for them to mimic. From left to right: mad, happy, sticking the tongue out, sad, and surprise.

database was classified as having European descent. Likewise, many of the subjects fell into the college (18-25) age group.

*EXPRESSIONS USED

Subjects were shown a series of expressions and asked to mimic the expressions. Four of the expressions can be categorized as some of Ekman's basic emotions: happy, sad, angry, and surprised [18]. The additional expression (the child with a tongue sticking out) was included because from a feature extraction perspective, it contains very unique and identifiable items (squinted eyes, furrowed brow, lips puckered in) [19].

The subjects were then asked to repeat two words from cue cards, 'business' and 'pleasure.' These expressions are an attempt to capture a different kind of dynamic from the face. Instead of detecting individuals based on emotional expressions, these dynamics are captured to show that individuals can also be identified simply through the way a person speaks.

DATA EXTRACTION

In order to create Active Appearance Model parameters, the face must be labeled according to a predefined outline. This outline acts as a representation of the object (in this experiment, face) with detailed representation of which points represent which facial features (for example, points 143 to 155 represent the mouth).



Figure 12: Example of a test subject’s expressions

STACKED ACTIVE SHAPE MODELS

In order to rapidly label over 300,000 face images, an automatic labeling system known as Stasm Active Shape Models, or Stasm [36]. Stasm was developed by Stephen Milborrow as a Master’s Thesis from the University of Cape Town, South Africa. Stasm acts as an extension of Cootes’ Active Shape Models by automatically locating the face and handling the iterations until convergence.

Like stated above, Cootes states that poor initial starting positions of the shape model can result in incorrect labeling of points. Stasm implements Rowley and Viola Jones techniques for face detection [48][55]. Once the face has been detected with either implementation, Stasm applies a trained shape model using the location of the face as the initial starting location. This helps lessen Cootes’ issue of poor initial location.

For this experiment, we took advantage of Stasm’s automated labeling system. Stasm allows the user to select between using Viola Jones (by default) or Rowley (not as accurate, but sometimes provides a better convergence). For the purposes of our experiment, we used only the default Viola Jones algorithm because it is considered the faster of the two algorithms in regard to Stasm’s processing. If, however, Viola-Jones was unable to find the face, the Rowley method was used. If both algorithms failed to located the face, it was manually labeled by hand. Stasm also allows two shape models to be ‘stacked’ on top of each other to increase chances at finding a more accurate location for points. This feature was not utilized during extraction because it increased processing time and using one shape model resulted in efficient point location.

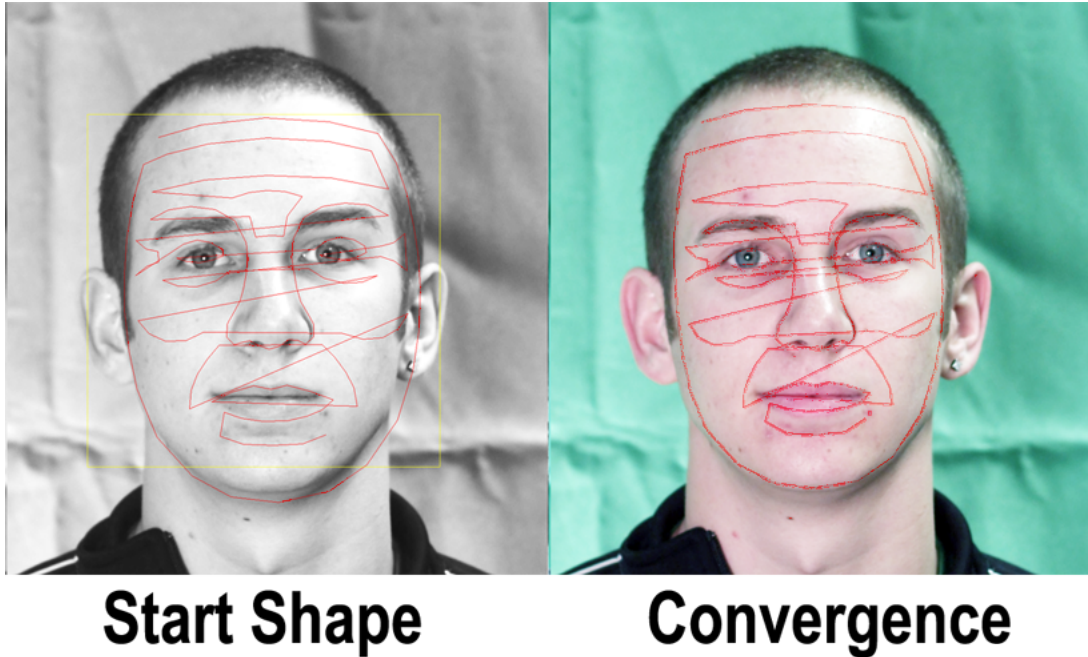


Figure 13: Example of Stasm's Initial Detection and Convergence

Before continuing with Active Appearance Models, we will discuss how exactly the Viola Jones and Rowley face detection algorithms work.

DETECTION OF THE FACE - VIOLA-JONES AND ROWLEY

The Viola-Jones face detection algorithm was developed by Paul Viola and Michael Jones at Cambridge that scans for features and edges of the face. In order to reduce computation time, the images are placed gray scale format (To make an image gray scale is to take the sum of the RGB values and replace them with the average of the values). The algorithm works under three kinds of features: a two, three, and four-rectangle feature system. The two-rectangle feature takes the sum of the pixels in the first rectangle are subtracted from the sum of the pixels in the second rectangle. The three-rectangle feature computes the sum of the two outer rectangles and subtracts them from the center rectangle. Finally, the four-rectangle combines two diagonal rectangles and subtracts them with the other two diagonal rectangles. If the difference between the sums is zero, there was no change in

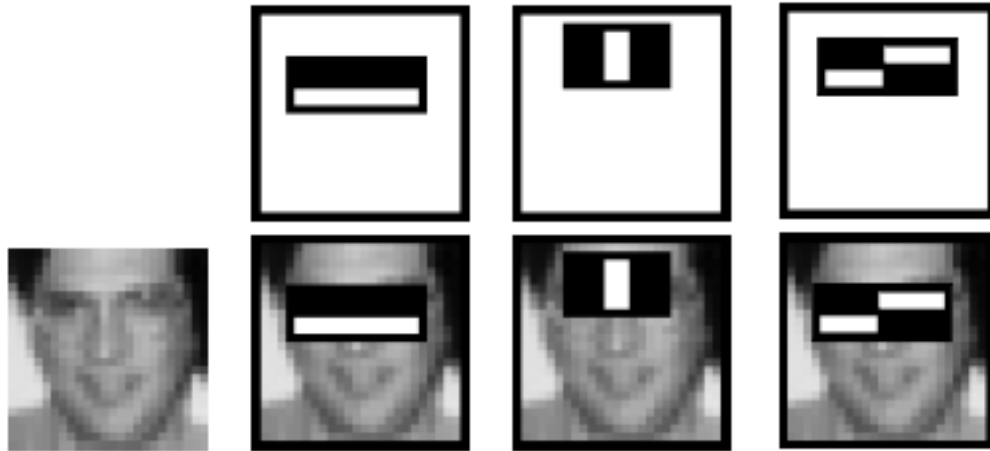


Figure 14: Viola-Jones employs edge detection from Haar-like feature vectors like above to detect the face.

the image's gradient; however, if the difference between the rectangles is beyond a given threshold, the algorithm treats that as locating a feature of the face[55].

The Rowley face detection algorithm was developed by Henry Rowley, Shumeet Baluja, and Takeo Kanade at Carnegie Mellon University as an attempt to locate upright, frontal faces in gray-scale images[48]. The Rowley algorithm is divided into two separate stages: first, it applies neural network-based filters to the image, and then uses an arbitrator to combine the outputs. These filters search the image for elements that might represent a face and then are merges the filters together to create a unified detection. Stasm initially searches for a face using Viola-Jones, however, if none can be found, it attempts to search again using Rowley's algorithm in order to minimize the chance of error[36].

DETECTION OF FACIAL FEATURES

Once Stasm has found the face via Viola-Jones or Rowley's algorithm it employes a 'whisker' system to find the features of the face. The 'whiskers' places orthogonal edge detectors along the landmark points. Once Stasm searches for the features of the face, the 'whiskers' will locate the difference in image intensity and contort the original shape model to fit the new edge[36].

Like many automatic feature extractors, some degree of training

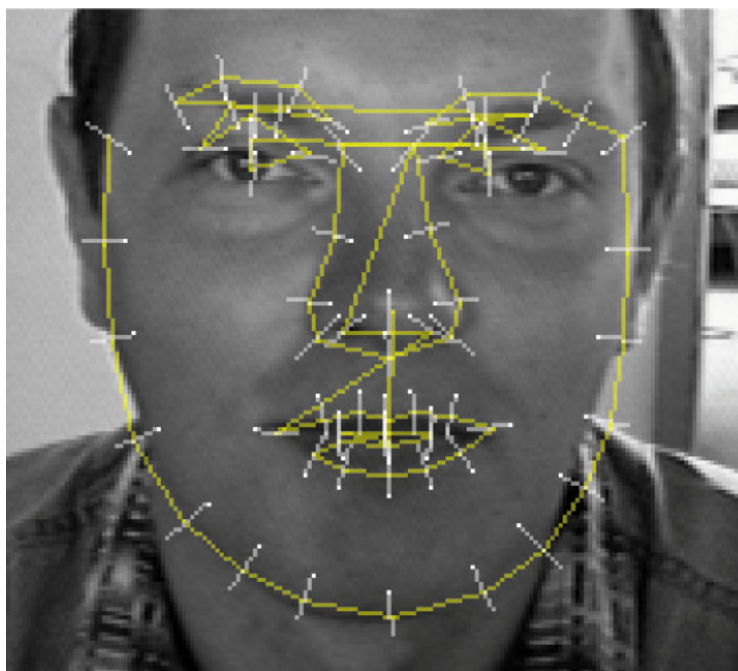


Figure 15: Example of the ‘whiskers’ used by Stasm. These whiskers are orthogonal to the shape model.

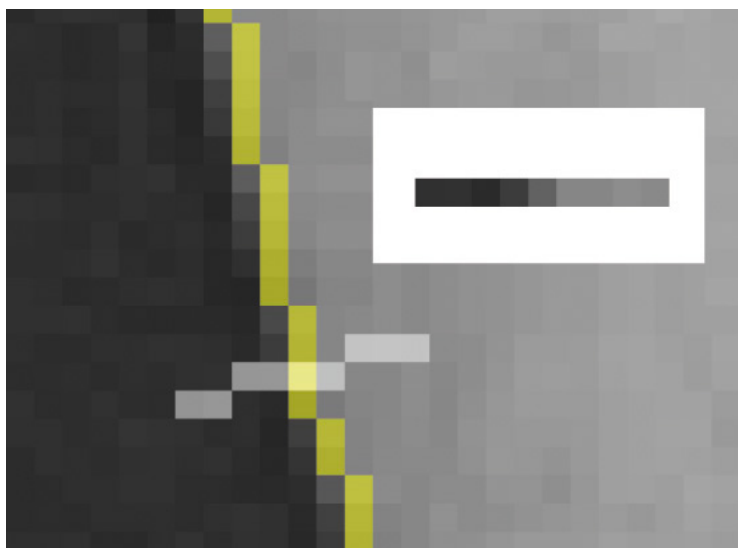


Figure 16: The ‘whiskers’ intersect with the shape model. Each intersection is centered at a small positive or negative displacement along the whisker, typically offset by 3 pixels.

was needed to create the shape model. Each subject was trained on the shape model with a neutral face, that is where no expressions were being mimicked. To ensure the shape model understood the variances of the mouth and eyes, some ‘happy’ and ‘mad’ expressions were included.

ACTIVE APPEARANCE MODELS

To show the ease of Active Appearance Models, Tim Cootes wrote a software package to use that labels images and conducts the principal component analysis and Delaunay Triangulation of the images[3]. The package includes a labeling application, various display tools, and a parameter generator. Due to the bulk number of images in the database, Cootes’ labeling software proved to be inefficient and Stasm was used instead. Cootes’ AAM training and parameter extracting applications, however, were used in order to obtain the appearance parameters of each frame. These frames were then condensed together and later sent through Python scripts to prepare them to be processed as features for Hidden Markov Models using HTK (discussed in the next section).

DATA ANALYSIS

HIDDEN MARKOV MODELS

Our research mirrors some techniques commonly used for speech recognition due to the success achieved using HMM [21][45][27]. The sequence of frames in an expression occur much like the phonetics in speech. In HMM, the various states that occur in speech recognition are reduced to a single state in face recognition. This is because unlike speech recognition, instead of changing from word to word, effectively changing state to state, the face of the test subject stays the same throughout the sequence and therefore, the state never changes states.

In forced alignment, a new recognition network is constructed for each input from a transcription and a dictionary. One of the main uses of forced alignment, however, is to determine the actual pronunciations used in the utterances used to train the HMM system to

$$d_t = \frac{\sum_{\theta=1}^{\Theta} \theta (c_{t+\theta} - c_{t-\theta})}{2 \sum_{\theta=1}^{\Theta} \theta^2}$$

Figure 17: A delta regression formula is used to calculate the difference between two frames

generate model level output transcriptions. Since the sequence of images of an expression only have one individual in the sequence, forced alignments allow for automatic recognition. While in speech recognition, forcing alignment is reserved for a bootstrapping process, taking the new networks and adding them to the training data.

Due to removing the forced alignment along with testing over various expressions, each subject has a greater degree of variance. In order to handle this variance, the number of states used in identification was considered a variable to tests. For the purpose of our experiment, the numbers of states were ranged from 5 to 15, in order to find an optimal number of states for recognition.

To create the models, the Hidden Markov Model Toolkit (HTK) was used. Originally created by the Machine Intelligence Laboratory of the Cambridge University Engineering Department before being acquired by Entropic Research Laboratory, which was later acquired by Microsoft. Although Microsoft owns the rights to HTK, it is considered open-source and Microsoft encourages further development of the toolset[57].

HTK is used to process the Viterbi and Baum-Welsh algorithms, along with the verification process. HTK was originally built for speech recognition, and therefore would not be able to handle visual feature vectors generated by Cootes' software. In order to convert appearance parameters to be usable by HTK, Python scripts extracted the data and condensed into parameter files that may be treated as traditional speech parameters but are actually AAM appearance parameters derived from the 100fps video.

FINDINGS

Each test was handled by the Hidden Markov Model Toolkit through six separate packages. The packages handled specific functions in processing the parameters for Markov models. All accuracy scores were processed through HTK’s internal ranking system; each score corresponds to the number of individual tests conducted versus the number correctly identified. In order to verify its identification, HTK looks at a predefined label. If the label was not present, HTK would just display its ‘best guess’. The goal of the tests is to show that the system can identify individuals. The delta-coefficient-based tests are the primary tests, as they handle face dynamics without using the face’s texture.

EXPERIMENTS

Testing was divided into six separate styles: static-only coefficients, static-delta coefficients, delta-only coefficients, delta-acceleration coefficients, combined delta coefficients, and single delta coefficients versus combined delta coefficients. Each style consisted of four tests, with the final scores being the average score of each test. The percentage of each test is based on the number of testing cases that were correctly identified against the total number of test cases in the test. The combined delta coefficients testing style contained 1,512 tests, divided evenly over 63 test subjects, 6 states, and 4 trials. The remaining testing styles consisted of 10,584 tests, divided evenly over 63 test subjects, 7 expressions, 6 states, and 4 trials. There were a total of 45,360 tests were conducted for this hypothesis.

STATIC COEFFICIENTS

In order to represent data in Active Appearance Models, parameter coefficients, or appearance parameters, are used to show the deviation from the mean face. The appearance parameters indicate the shape and texture-levels of the model. In a traditional face recognition system, identifying individuals on a single image, the appearance parameters would be sufficient enough to discern each individual[54][14]. The

appearance parameters are referred to as the static coefficients in the testing process since they represent the data of each frame in the video as is. In an attempt to recognize individuals off a non-neutral face (one holding an expression) Cootes found an average score of 88% accuracy over 400 images - 200 trained and 200 tested across 20 individuals[14]. This score is mirrored in the static coefficient tests, holding a similar average accuracy of 90% with the highest score being 100% accuracy (found in 16 tests, see Fig. 18 for individual scores) and the lowest being 69% (found in 3 tests). The static coefficient tests act as a baseline score towards the rest of the tests. Since the static coefficients are the unchanged appearance parameters, the tests act as a traditional face recognition system.

STATIC-DELTA COEFFICIENTS

Delta coefficients refer to the difference in static coefficients over the current frame and the frame immediately before. These coefficients are also commonly known as velocity, or the rate of change over a given parameter from one image frame and the last previous frame [See Fig. 17 for the regression formula]. The delta coefficients can help distinguish individuals based on how they move their face, without the need of the appearance parameters. When the delta coefficients were included in the static coefficient tests, there was no sign of difference between the two tests. The experiment mirrored the static-only results with an average accuracy of 90% with the highest score being 100% accuracy and the lowest being 69% (these scores directly correspond with Static-Only tests). Since the static coefficients are the appearance parameters, the unique movements of the face did not improve the scores.

DELTA-ONLY COEFFICIENTS

In order to test accuracy using only facial dynamics, the static coefficients were removed from the testing data in order to test based only on the movements of the face. As stated earlier, appearance models allow for a greater degree of detail in the textures of the face [4]. This allows the delta coefficients the ability to distinguish the formation of

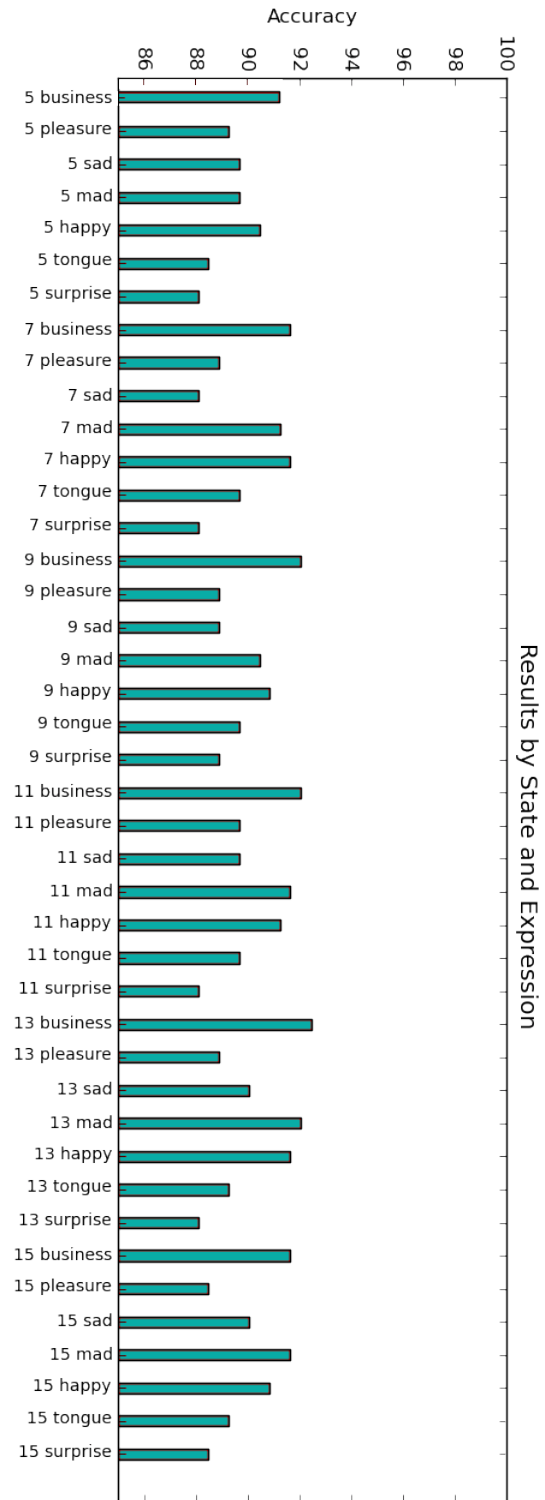


Figure 18: Results of Static-Only Coefficients Testing based and State and Expression

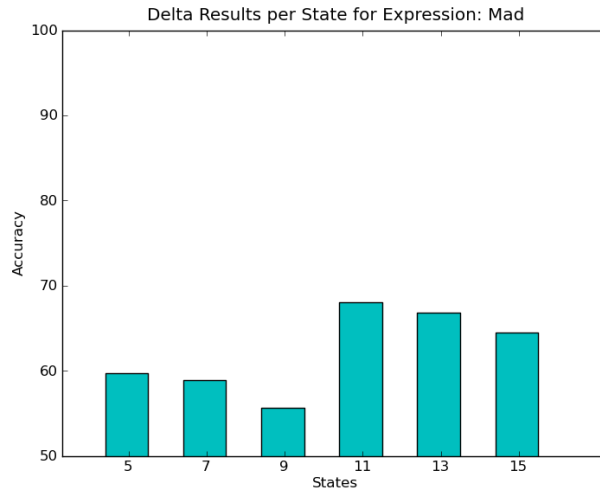


Figure 19: Results of Delta-Only ‘Mad’ Tests aligned by States

a given texture in the face, such as a dimple or wrinkles. The delta coefficients were able to generate an average accuracy of 76% successful recognitions. The tests involving solid facial expressions (mad, happy, stick-out the tongue, sad, and surprise) generated an average accuracy of 77% with the best scores reaching 100% (found once in ‘surprise’ test at 9 states) and the worst hitting 31% (found at state 15 on expression ‘sticking out tongue’). This worst score was during the 15 states model. The 15 state model overall showed a sharp decline in scores compared to the 13 state models (overall accuracy of 13 states was 90% while 15 states was 70%). The lower scores possibility indicates a degree of over training or training on redudant dynamic areas that were shown in multiple subjects. The remaining tests were in order to test accuracy against how an individual says a given word in a neutral state. The average accuracy of these tests was 68%. This is one of the most significant results of the these experiments, as it demonstrates accuracy rates comparable to other biometric results [54, 14, 42, 45, 8, 58] using only velocity parameter measurements, a pure face dynamic.

DELTA-ACCELERATION COEFFICIENTS

Acceleration coefficients are created using the delta regression formula [Fig. 17] on the delta coefficients. In relation to facial dynamics, this represents the acceleration at which the textures are changing

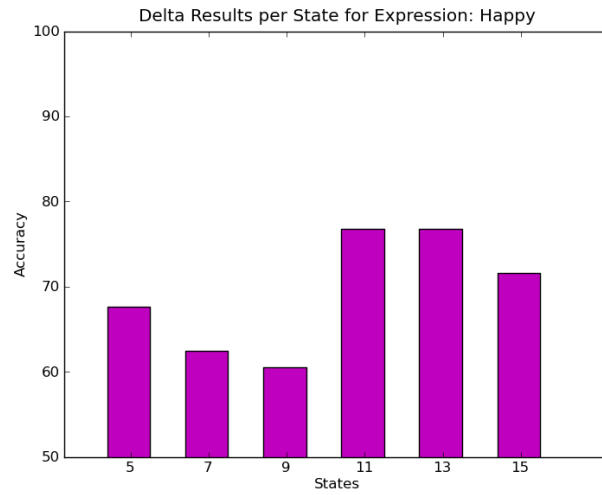


Figure 20: Results of Delta-Only 'Happy' Tests aligned by States

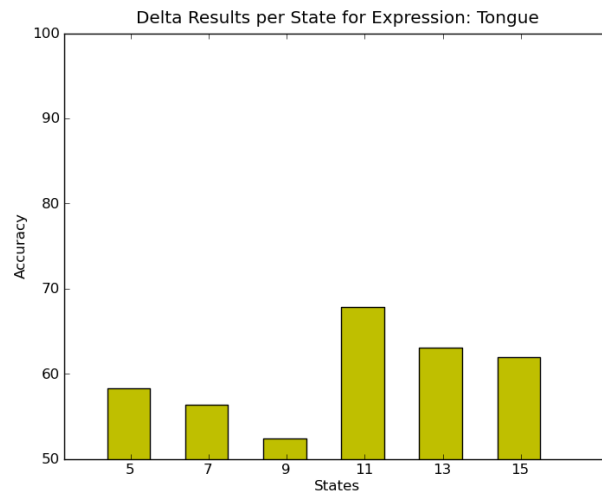


Figure 21: Results of Delta-Only 'Sticking-Out Tongue' Tests aligned by States

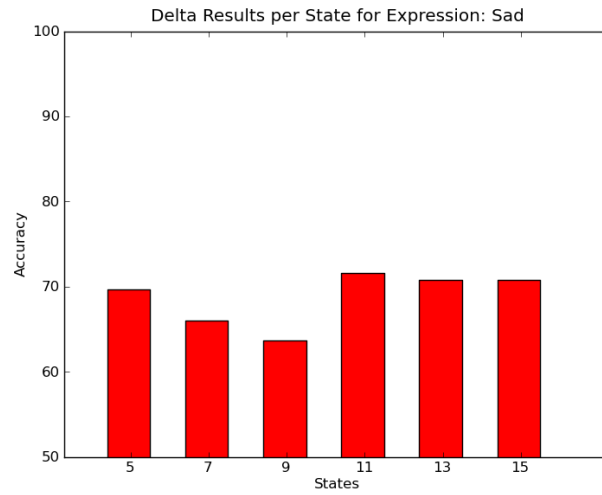


Figure 22: Results of Delta-Only ‘Sad’ Tests aligned by States

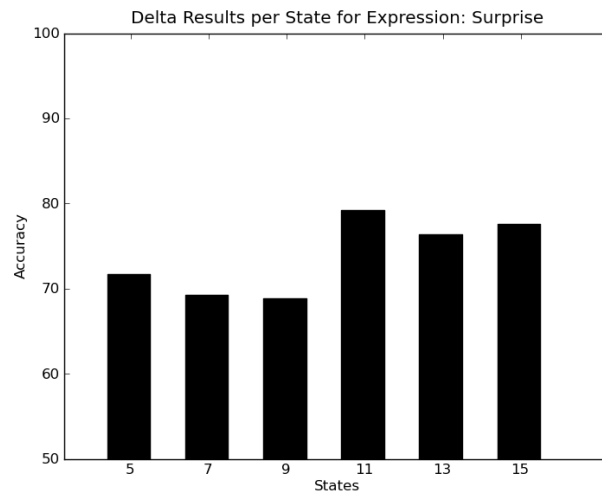


Figure 23: Results of Delta-Only ‘Surprise’ Tests aligned by States

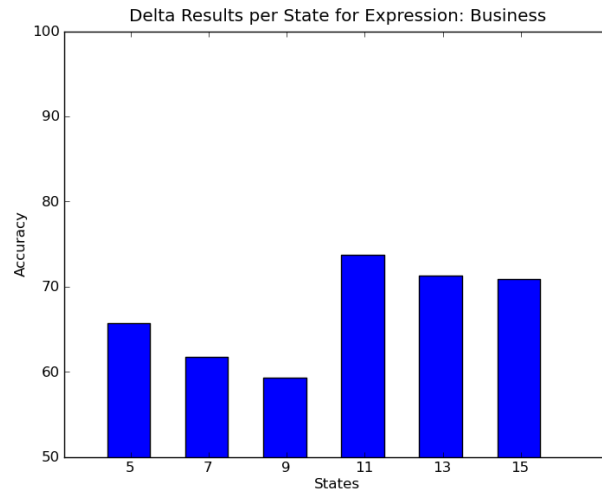


Figure 24: Results of Delta-Only ‘Business’ Tests aligned by States

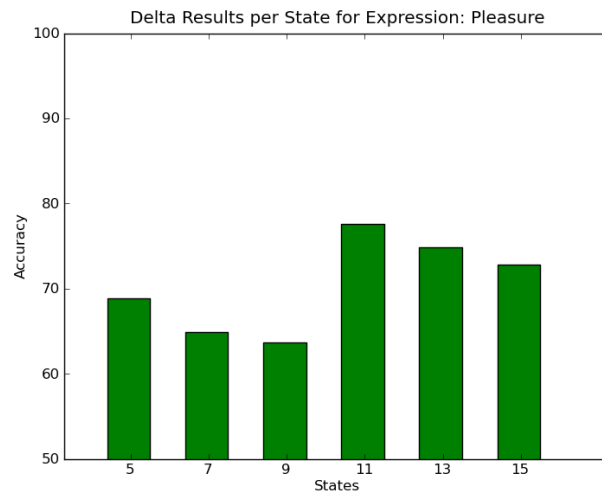


Figure 25: Results of Delta-Only ‘Pleasure’ Tests aligned by States

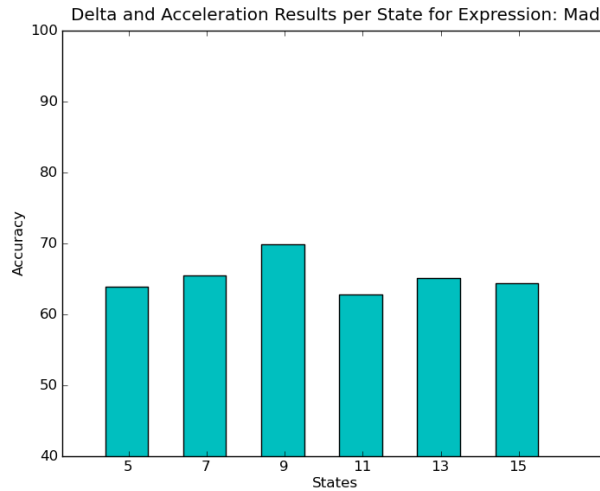


Figure 26: Results of Delta-Acceleration ‘Mad’ Tests aligned by States

on the original appearance parameters. Similar to adding delta coefficients to the static parameters, the acceleration coefficients were added to the delta in order to increase accuracy. Unlike the static-delta tests, the delta-acceleration tests showed a decline in accuracy to an average of 60% with the highest score reaching 80% (found in test ‘mad’ at state 15) and lowest being 35% (found at state 15 with expression ‘business’). This is an interesting result, because when delta coefficients were added to static coefficients, the scores stayed the same. The expected result was that the delta-acceleration tests would slightly increase scores or at least stay even with the delta-only test. In Fox’s speaker identification paper[22], a larger number of states were needed to appropriately score delta-acceleration only tests. While tests conducted in these experiments stop with 15 states, Fox’s experiments expended to 30 states.

COMBINED DELTA COEFFICIENTS

The previous experiments separated each test by one of the seven expressions the subjects were asked to mimic. Since the previous tests were dependent on the expression being known, for the combined delta coefficient tests, the delta parameters were joined together sequentially and tested against each other. In this test, it allows the system to iden-

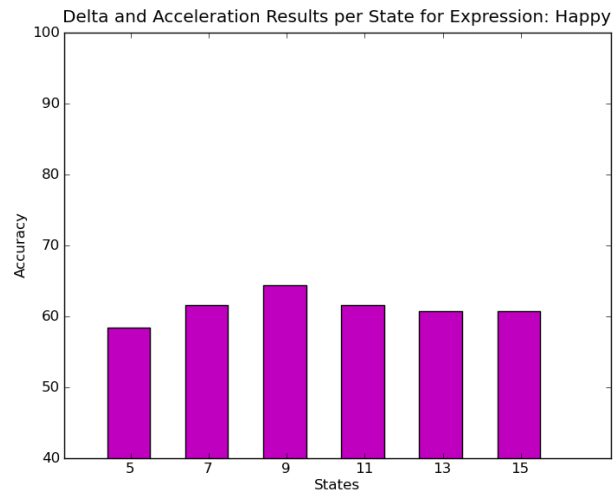


Figure 27: Results of Delta-Acceleration ‘Happy’ Tests aligned by States

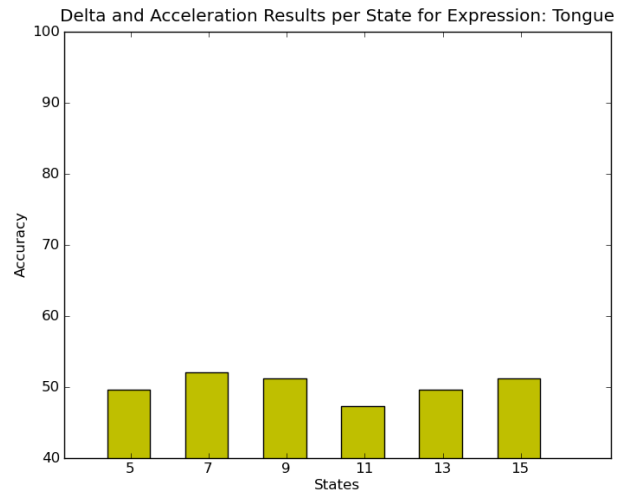


Figure 28: Results of Delta-Acceleration ‘Sticking-Out Tongue’ Tests aligned by States

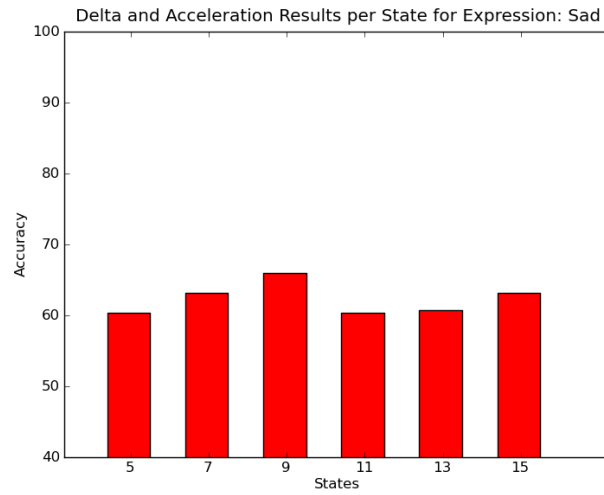


Figure 29: Results of Delta-Acceleration ‘Sad’ Tests aligned by States

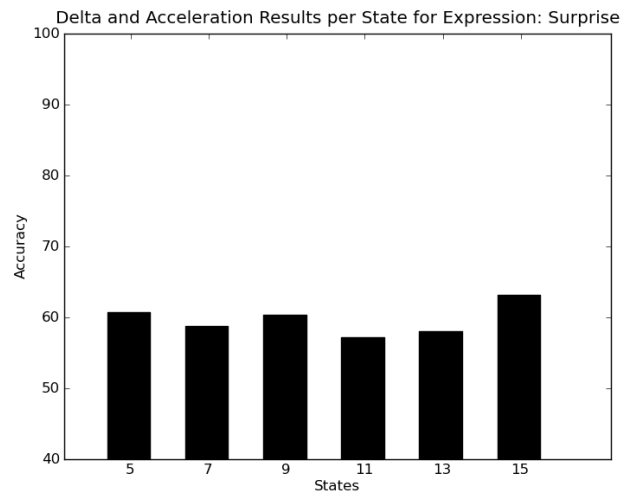


Figure 30: Results of Delta-Acceleration ‘Surprise’ Tests aligned by States

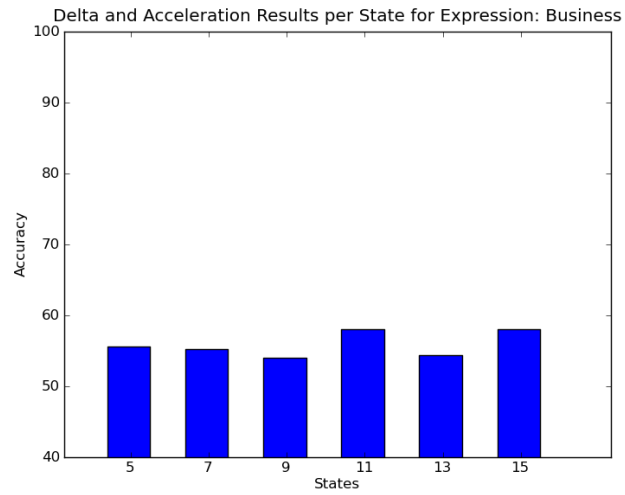


Figure 31: Results of Delta-Acceleration ‘Business’ Tests aligned by States

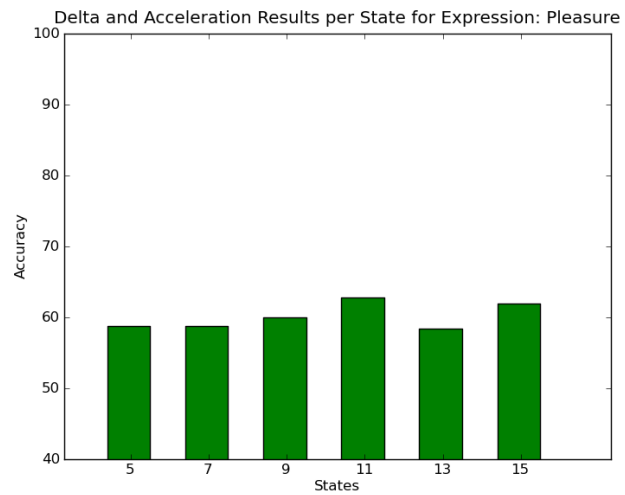


Figure 32: Results of Delta-Acceleration ‘Pleasure’ Tests aligned by States

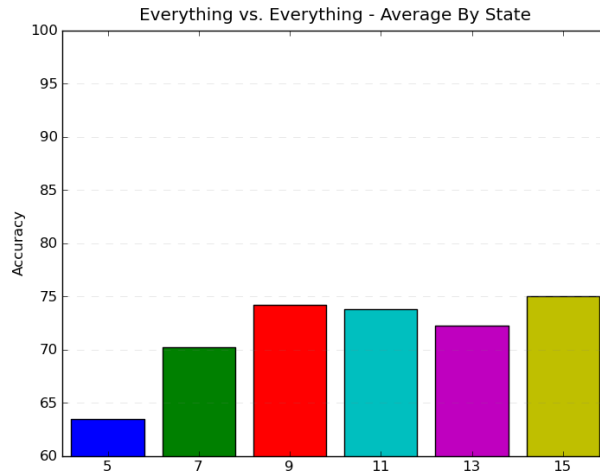


Figure 33: Results of Combined Delta Coefficients sorted by State

tify individuals based on a sequence of events. The average score was 71% with definitive score increases when the number of states were introduced. When the Markov models were built with 5 states, the average score of all tests was below 65% (the lowest being 47%); however, upon increasing the number of states to 7 and again to 9 showed significant increases in accuracy, with 7 states having the maximum score of 85%. After 9 states, the scores began to normalize. Interestingly, 9 states would exhibit lower score results when the expressions were predefined; however shows the most increase when the parameters are combined.

SINGLE EXPRESSION DELTA COEFFICIENTS VERSUS COMBINED DELTA COEFFICIENTS

The final test involved training the Markov models on all combined delta parameters, and then testing the model against individual expressions. This can help reduce feature extraction processes by allowing the system be trained without having to divide the subject's mannerisms. For testing, any captured expression can be compared against the entire model. As a result, there was a 4% increase in the average scoring (75%) compared to testing with combined delta coefficients. Like the previous combined delta test, as the number of states the model was built on increased, the level of accuracy also increased,

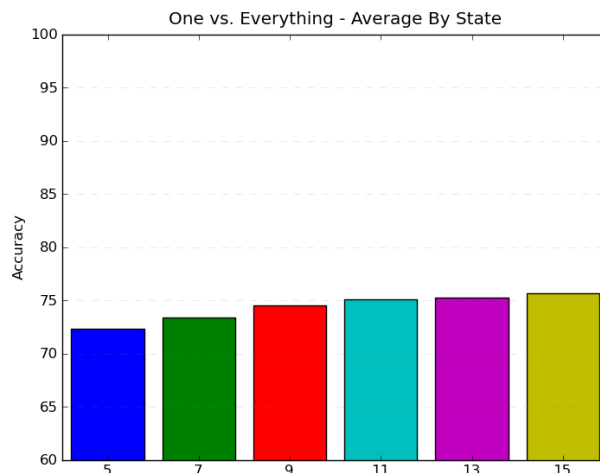


Figure 34: Results of Single Expression Delta Coefficients Combined Delta Coefficients sorted by State

although with diminishing returns.

OBSERVATIONS

In each test, increasing the number of states allowed for higher levels of accuracy, before evening out. After 11 states, the overall benefit in state increases drops to marginal improvements in combined expression experiments. In the separated expression experiments, scores began to drop during 15 state models, potentially due to over training. Further investigation of the number of states used should be pursued in order to see if the Markov models had plateaued.

Additionally, the spoken words showed equal scores compared to the expressions. This can help identify subtleties within the face as small as the way a person says a word can be used as a means to identify them. While not exactly a microexpression, it paves the way for studies on actual microexpression research.

	Five	Seven	Nine	Eleven	Thirteen	Fifteen
Business	65.7125	61.73	59.3325	73.6825	71.2775	70.8925
Pleasure	68.8775	64.9025	63.705	77.6375	74.86	72.8625
Sad	69.6225	66.0275	63.645	71.6075	70.8125	70.8075
Mad	59.6775	58.865	55.685	68.06	66.835	64.44
Happy	67.665	62.475	60.495	76.8025	76.8175	71.6425
Tongue	58.3325	56.35	52.38	67.8575	63.095	61.905
Surprise	71.6475	69.2525	68.845	79.21	76.41	77.605

Table 1: Average Scores Across Each Expression and State Combination for Delta-Only Coefficients

FUTURE WORK

There is still a largely unexplored area of biometrics involving facial dynamics. Since the thesis explores the possibility of using facial dynamics, it can become the basis towards exploring other expressions for recognition. The expressions used in the experiment were mostly basic emotions described by psychologists and philosophers [11, 51, 25, 49, 18]. Unfortunately, emotions are not a static state, where one is only one emotional state at a time; there is a grey area of mixed emotions that may or may not have physical representation in the face [7, 44].

One of the drawbacks of the thesis is that the experiment cued subjects to the expressions they were supposed to create, thus creating false expressions with no emotional attachment. While eliciting genuine expressions is still a difficult task, it could hold information over more subtle differences in facial expressions. This would help support Paul Ekman’s microexpressions and ‘leakages’ claim. If a ‘leakage’ can be identified, it would require a system that would be able to note the microexpressions of the face during a genuine deception expression. Monitoring genuine microexpressions would be able to show, if any, the minor changes in the face in some of the more complex emotional states. Future work could involve applying a system described here to those other areas of expression research, for both anthropological and computational experiments.

This thesis focused on basic emotional expressions; however, with the spoken word expressions, the thesis shows that there are differences

in something as subtle as speaking. In society, there are a countless number of words and phonemes that could be classified for the purpose of computerized lip reading[37]. By identifying the subtleties of the expressions involved in speaking, a lip reading system could be developed. Instead of focusing on specific words, investigation could begin to detect phonemes pronounced, which could then translate into words.

CONCLUSION

Humans have shown that they do not need the actual face in order to identify people they know. They are able to use gestures and expressions that are unique to the individual known as gait analysis. In this thesis, a form similar to gait analysis was using the dynamics of the face in order to train a computer to be able to identify individuals. Security protocols that are dependent on ensuring a person's identification can use gait analysis to reduce the change of a false-positive matching, despite correct facial identification.

By filming tests subjects at 100 frames per second, the data extracted for each subject allowed for higher quality textures and more subtle movements to stand out per frames. With a higher frame rate, the need for an automated feature labeling system was an important role that Stasm filled successfully. With Stasm's accuracy close to that of a human labeler, instead of taking years to manually label over 400,000 images, the process was completed in roughly 3 days. Active Appearance Modeling used Principal Component Analysis in order to take the average face shape and texture. Now that an average model existed, the appearance parameters of each frame could represent a particular face and its pose in that frame. These parameters were then passed through a delta regression formula, or used as is, to create the static-, delta-, and acceleration-coefficients that were analyzed using Hidden Markov Models. As the states increased, the accuracy in verification increased with diminishing returns. The static coefficients were able to achieve an average 90% accuracy, showing that the textures of the face were a sufficient identifier when not dealing with non-optimal capturing environments. Delta coefficients were able to score an average of 75%, although did not show and increase when the acceleration-coefficients were added to training, potentially due to a low number of states.

The hypothesis of this experiment was to demonstrate that a computer system could identify individuals after being trained on the dynamics of the face as expressions were elicited. By achieving scores ranging from 68% to 90%, accuracy levels similar to other research

biometrics [54, 14], this hypothesis is shown to be correct. Future work in this area could produce very useful biometric and expression-based technologies.

REFERENCES

- [1] Cristinacce D. Taylor C. Caunce, A. and T. Cootes. Locating facial features and pose estimation using a 3d shape model. In *ISVC '09: Proceedings of the 5th International Symposium on Advances in Visual Computing*, pages 750–761, Berlin, Heidelberg, 2009. Springer-Verlag.
- [2] T. F. Cootes, D. H. Cooper, C. J. Taylor, and J. Graham. Trainable method of parametric shape description. *Image Vision Comput.*, 10(5):289–294, 1992.
- [3] T.F. Cootes. Statistical models of appearance for computer vision. tim cootes home page, March 2009. <http://personalpages.manchester.ac.uk/staff/timothy.f.cootes>.
- [4] T.F. Cootes, G. Edwards, and C.J. Taylor. Comparing active shape models with active appearance models. In *British Machine Vision Conference*, volume 1, pages 173–183, 1999.
- [5] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *Computer Vision ECCV98*, page 484, 1998.
- [6] T.F. Cootes and C.J. Taylor. Active shape models - ‘smart snakes’. In *Proceedings of the British Machine Vision Conference*, 1992.
- [7] T. Dalgleish and M. Power. *Handbook of cognition and emotion*. Wiley Online Library, 1999.
- [8] I.G. Damousis, D. Tzovaras, and E. Bekiaris. Unobtrusive multimodal biometric authentication: The humabio project concept. *EURASIP Journal on Advances in Signal Processing*, 2008:1–11, 2008.
- [9] Charles Darwin. *Expression of the Emotions in Man and Animals, The*. Oxford University Press Inc, 2002.
- [10] L Delaunay. Triangulation for planar graphs. *Discrete and Computational Geometry*, 44:1–29, 1988.

- [11] R. Descartes and S. Voss. *The passions of the soul*. Hackett Publishing Company, 1989.
- [12] G.B. Duchenne and R.A. Cuthbertson. *The mechanism of human facial expression*. Cambridge University Press, 1990.
- [13] S.R. Eddy. Hidden markov models. *Current Opinion in Structural Biology*, 6(3):361–365, 1996.
- [14] Cootes T.F. Edwards, G.J. and C.J. Taylor. Face recognition using active appearance models. *Computer Vision ECCV98*, page 581, 1998.
- [15] P. Ekman. *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Owl Books, 2004.
- [16] P. Ekman. *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. WW Norton & Company, 2009.
- [17] P. Ekman et al. Universals and cultural differences in facial expressions of emotion. In *Nebraska symposium on motivation*, volume 19, pages 207–283. University of Nebraska Press, 1971.
- [18] P. Ekman and W.V. Friesen. *Unmasking the face: A guide to recognizing emotions from facial clues*. Prentice-Hall Englewood Cliffs, NJ, 1975.
- [19] P. Ekman, W.V. Friesen, and J.C. Hager. *Facial action coding system*. Consulting Psychologists Press Palo Alto, CA, 1978.
- [20] Marcos Faundez-Zanuy and Joan Fabregas. On the relevance of facial expressions for biometric recognition. In *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction: COST Action 2102 International Conference, Patras, Greece, October 29-31, 2007. Revised Papers*, pages 32–43, Berlin, Heidelberg, 2008. Springer-Verlag.
- [21] Cohn J.J. Gross R. Fox, N. and R. Reilly. Person identification using automatic integration of speech, lip, and face experts. In *ACM SIGMM 2003 Multimedia Biometrics Methods and Applications Workshop (WBMA03)*, pages 25–32. ACM Press, 2003.

- [22] OMullane B. Fox, N. and R. Reilly. Audio-visual speaker identification via adaptive fusion using reliability estimates of both modalities. In *Audio-and Video-Based Biometric Person Authentication*, pages 787–796. Springer, 2005.
- [23] D. Gafurov. A survey of biometric gait recognition: Approaches, security and challenges. In *Nik-2007 Conference*, 2007.
- [24] H. Hill and A. Johnston. Categorizing sex and identity from the biological motion of faces. *Current Biology*, 11(11):880–885, 2001.
- [25] T. Hobbes and J.C.A. Gaskin. *Leviathan*. Oxford University Press, USA, 1998.
- [26] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3-4):321, 1936.
- [27] M. Hwang and X. Huang. Shared-distribution hidden markov models for speech recognition, 1991.
- [28] A. Jain, S. Prabhakar, and S. Pankanti. Twin test: On discriminability of fingerprints. In *Audio-and Video-Based Biometric Person Authentication*, pages 211–217. Springer, 2001.
- [29] B. Knight and A. Johnston. The role of movement in face recognition. *Visual Cognition*, 4(3):265–273, 1997.
- [30] A.W.K. Kong, D. Zhang, and G. Lu. A study of identical twins’ palmprints for personal verification. *Pattern Recognition*, 39(11):2149–2156, 2006.
- [31] A. Lanitis, C.J. Taylor, and T.F. Cootes. Toward automatic simulation of aging effects on face images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):442–455, 2002.
- [32] L. Lee and W. Grimson. Gait analysis for recognition and classification. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 148–155. IEEE, 2002.

- [33] X. Liu and T. Chen. Video-based face recognition using adaptive hidden markov models. 2003.
- [34] M. Loeve. Probability theory, vol. ii. *Graduate Texts in Mathematics*, 46:0–387.
- [35] Thacker N.A. Luettin, J. and S.W. Beet. Speaker identification by lipreading. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 1, pages 62–65. IEEE, 2002.
- [36] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model, 2008.
- [37] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern recognition letters*, 17(2):155–162, 1996.
- [38] Y. Nakata and M. Ando. Lipreading method using color extraction method and eigenspace technique. *Systems and Computers in Japan*, 35(3):12–23, 2004.
- [39] A.J. O’Toole, D. Roark, and H. Abdi. Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, 6:262–266, 2002.
- [40] Harms J. Snow S. L. Hurst D. R. Pappas M. R. O’Toole, A.J. and H. Abdi. A video database of moving faces and people. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):812–816, 2005.
- [41] 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.
- [42] Gurbuz S. Tufekci Z. Patterson, E.K. and J.N. Gowdy. Moving-talker, speaker-independent feature study, and baseline results using the cuave multimodal speech corpus. *EURASIP Journal on Applied Signal Processing*, 2002(1):1201, 2002.

- [43] K. Pearson. Liii. on lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572, 1901.
- [44] R. Plutchik. The nature of emotions. *American Scientist*, 89:344, 2001.
- [45] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [46] RED Digital Cinema, 20291 Valencia Circle, Lake Forest, CA, USA 92630. *RED ONE Operation Guide Build*.
- [47] M. Roberts, T. Cootes, E. Pacheco, T. Oh, and J. Adams. Segmentation of lumbar vertebrae using part-based graphs and active appearance models. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pages 1017–1024, 2009.
- [48] H.A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(1):23–38, jan. 1998.
- [49] M.E.P. Seligman. *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. Free Pr, 2002.
- [50] L.I. Smith. A tutorial on principal components analysis. *Cornell University, USA*, 51:52, 2002.
- [51] B. Spinoza. Ethics, trans. *W. Hale White, fourth edn (Oxford: Oxford University Press)*, 1910.
- [52] Z. Sun, A.A. Paulino, J. Feng, Z. Chai, T. Tan, and A.K. Jain. A study of multibiometric traits of identical twins. 2010.
- [53] M. Tistarelli, M. Bicego, and E. Grosso. Dynamic face recognition: From human to machine vision. *Image and Vision Computing*, 27(3):222–232, 2009.

- [54] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.
- [55] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. volume 1, pages I–511 – I–518 vol.1, 2001.
- [56] J.H. Warfel and J.C. Posnick. The head, neck and trunk. *Plastic and Reconstructive Surgery*, 92(6):1196, 1993.
- [57] S.J. Young. The htk hidden markov model toolkit: Design and philosophy. *Entropic Cambridge Research Laboratory, Ltd*, 2:2–44, 1994.
- [58] T. Zhang, X. Li, D. Tao, and J. Yang. Multimodal biometrics using geometry preserving projections. *Pattern Recognition*, 41(3):805–813, 2008.