UNIVERSITY OF DENVER

A MULTI-MODAL APPROACH FOR FACE MODELING AND RECOGNITION

By

Huanghao Feng

A COMPREHENSIVE EXAM

Submitted to the Faculty
of the University of Denver
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Denver, Colorado

December 2007

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

A MULTI-MODAL APPROACH FOR FACE MODELING AND RECOGNITION

Mohammad Hossein Mahoor

Approved:

Dr. Mohamed Abdel-Mottaleb Associate Professor of Electrical and Computer Engineering	Dr. Terri A. Scandura Dean of Graduate School
Dr. Shahriar Negahdaripour Professor of Electrical and Computer Engineering	Dr. Kamal Premaratne Professor of Electrical and Computer Engineering
Dr. James W. Modestino Professor of Electrical and Computer Engineering	Dr. Ahmad El-gammal Assistant Professor of Computer Science Rutgers University

FENG, HUANGHAO (Ph.D., Electrical and Computer Engineering)
(December 2007)

Abstract of a comprehensive exam at the University of Denver.

Dissertation supervised by Dr. Mohammad Mahoor.
No. of pages in text 39.

This dissertation describes a new methodology for multi-modal.

To my beloved mother, deceased father,

and

to my lovely wife.

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude to Professor Abdel-Mottaleb, who introduced me to the field of computer vision, and helped me to jump start my research career in this field. He encouraged me to pursue novel ideas and provided me with exceptional experience and knowledge.

My thanks also go to the other members of the committee, Dr. Shahriar Negahdaripour, Dr. James W. Modestino, Dr. Kamal Premaratne, and Dr. Ahmad El-gammal for their valuable comments and suggestions. In particular, I want to thank Dr. Negahdaripour for his valuable support, comments and encouragement during my Ph.D study.

More than everyone, I indebted to my mother for her enthusiastic encouragement, prayers and unlimited support during all stages of my life. Also, I must express my appreciation to my beloved wife, Eshrat, my brothers and sisters.

I would like to extend my thanks to the faculty and staff members of the Electrical and Computer Engineering Department at the University of Miami, especially Ms. Clarisa Alvarez and Ms. Rosamund Coutts for their role in my education and various resources made available to me to do my research.

Also, I would like to thank my fellow lab members, specially, Dr. Nasser Al-Ansari, Dr. Pezhman Firoozfam, Dr. Omaima Nomir, Dr. Charay Leurdwick, Hamed Pirsiavash, Feng Niu, Ali Taatian, Steven Cadavic, Jindan Zhou, Hossein Madjidi,

Behzad M. Dogahe, Dr. Hongsheng Zhang, Muhammad Rushdi, and Dr. Nuno Gracias for their collaboration during the completion of this thesis.

In addition, I want to thank Dr. Tapia, Dr. Asfour, and Mr. Ali Habashi for their support and help during my study at the University of Miami.

Contents

List of Figures

List of Tables

List of Acronyms

2-D: Two Dimensional

3-D: Three Dimensional

ARG: Attributed Relational Graph

ASM: Active Shape Model

CMC: Cumulative Match Characteristic

EBGM: Elastic Bunch Graph Matching

FDA: Fisher Discriminant Analysis

FM: False Match

FMR: False Match Rate

FNM: False Non Match

FNMR: False Non Match Rate

FRGC: Face Recognition Grand Challenge

FRVT: Face Recognition Vendor Tests

HD: Hausdorff distance

ICP: Iterative Closest Points

LTS-HD: Least Trimmed Square-HD

MSE: Mean Square Error

PCA: Principal Component Analysis

ROC: Receiver Operating Characteristic

SFFS: Sequential Floating Forward Selection

Chapter 1

Introduction

1.1 Autism Spectrum Disorders (ASD)

Autism is a general term used to describe a spectrum of complex developmental brain disorders causing qualitative impairments in social interaction and results in repetitive and stereotyped behaviors. Currently one in every 88 children in the United States are diagnosed with ASD and government statistics suggest the prevalence rate of ASD is increasing 10-17 percent annually [9]. Children with ASD experience deficits in appropriate verbal and nonverbal communication skills including motor control, emotional facial expressions, and eye gaze attention [10]. Currently, clinical work such as Applied Behavior Analysis (ABA) [11] [12] has focused on teaching individuals with ASD appropriate social skills in an effort to make them more successful in social situations [1]. With the concern of the growing number of children diagnosed with ASD, there is a high demand for finding alternative solutions such as innovative computer technologies and/or robotics to facilitate autism therapy. Therefore,

research into how to design and use modern technology that would result in clinically robust methodologies for autism intervention is vital. In social human interaction, non-verbal facial behaviors (e.g. facial expressions, gaze direction, and head pose orientation, etc.) convey important information between individuals. For instance, during an interactive conversation, the peer may regulate their facial activities and gaze directions actively to indicate the interests or boredom. However, the majority of individuals with ASD show the lack of exploiting and understanding these cues to communicate with others. These limiting factors have made crucial difficulties for individuals with ASD to illustrate their emotions, feelings and also interact with other human beings. Studies have shown that individuals with autism are much interested to interact with machines (e.g. computers, iPad, robots, etc.) than humans [6]. In this regard, in the last decade several studies have been conducted to employ machines in therapy sessions and examine the behavioral responses of people with autism. These studies have assisted researchers to better understand, model and improve the social skills of individuals on the autism spectrum. This thesis presents the methodology and results of a study that aimed to design a humanoid-robot therapy sessions for capturing, modeling and enhancing the social skills of children with Autism. In particular we mainly focus on gaze direction and joint attention modeling and analysis and investigate how the ASD and Typically Developing (TD) children employ their gaze for interacting with the robot. In the following section, we have a brief introduction of the existing assistive robots in the following section and how they have been used in autism applications.

1.2 Socially Assistive Robotics

Socially Assistive Robotics (SAR) can be considered as the intersection of Assistive Robotics (AR) and Socially Interactive Robotics (SIR), which has referred to robots that assist human with physical deficits and also can provide certain terms of social interaction abilities [5]. SAR contains all properties of SIR described in [6], and also a few additional attributes such as: 1) user populations (different groups of users, i.e. elders; individuals with physical impairments; kids diagnosed with ASD; students); 2) social skills (i.e. speech ability; gestures movement); 3) objective tasks (i.e. tutoring; physical therapy; daily life assistance); 4) role of the robot (depends on the task the robot has been assigned for) [5]. Companion robots [7] is one type of SAR that are widely used for elderly people for health care supports. Research shows that this type of social robots can reduce stress and depression of individuals in elderly stage [8]. Service social robots are able to accomplish a variety of tasks for individuals with physical impairments [9]. Studies have shown that SAR can be used in therapy sessions for those individuals who suffer from cognitive and behavioral disorders (e.g. Autism). SAR provides an efficient helpful medium to teach certain types of skills to these groups of individuals [10] [11] [12]. Nowadays, there are very few companies that have been designing and producing socially assistive robots. The majority of existing SARs are not commercialized yet and because of being expensive and not well-designed user interfaces, they are mostly used forthe research purposes. Honda, Aldebaran Robotics and Hanson Robokind are the top leading companies that are currently producing humanoid robots. Ideally socially assistive robots can have

fully automated systems to detect and express social behaviors while interacting with humans. Some of the existing robot-human interfaces are semi-autonomous and they can recognize some basic biometrics (e.g. visual and audio commands of the user) and behavioral response. Besides, the majority of existing robots are very complicated to work with. Therefore in the last couple of years several companies have started to make these robots more user-friendly and responsive to both the user need and the potential caregiver commands [5]. Intelligent SARs aim to have the capability to recognize visual or audio commands, objects, and specific human gestures. Some of these robots have the ability of detect human face or basic facial expressions. For instance, ASIMO, a robot developed by Honda, it has a sensor for detecting the movements of multiple objects by using visual information captured from two cameras on its head. Plus its "eyes" can measure the distance of the objects from the robot [13]. Another example is from Aldebaran Robotics which designs small size humanoid robots, called NAO. NAO robot has two cameras attached that are used to capture single images and video sequences. This capturing module enables NAO to see the different sides of an object and recognize it for future use. Furthermore, NAO has a remarkable capability of recognizing faces and detecting moving objects. Both of the aforementioned robots have speech recognition system. They can interpret voice commands to accomplish a certain set of tasks which have been pre-programmed in the system. NAO is able to identify words for running specific commands. However ASIMO is able to distinguish between voices and other sounds. This feature empowers ASIMO to perceive the direction of human's speaker or recognize other companion robots by tracking their voice [14]. These robots can also speak in many different

languages. For example, NAO can speak in English, French, Chinese, Japanese and other languages up to more than ten languages. This feature gives the robot a great social communication functionality to interact with humans from all over the world.

1.2.1 Socially Assistive Robots for Autism Therapy

Socially assistive robots are emerging technologies in the field of robotics that aim to utilize social robots to increase engagement of users as communicating with robots, and elicit novel social behaviors through their interaction. One of the goal in SAR is to use social robots either individually or in conjunction with caregivers to improve social skills of individuals who have social behavioral deficits. One of the early applications of SAR is autism rehabilitation. As mentioned before, autism is a spectrum of complex developmental brain disorders causing qualitative impairments in social interaction. Children with ASD experience deficits in appropriate verbal and nonverbal communication skills including motor control, emotional facial expressions, and gaze regulation. These skill deficits often pose problems in the individual's ability to establish and maintain social relationships and may lead to anxiety surrounding social contexts and behaviors [1]. Unfortunately there is no single accepted intervention, treatment, or known cure for individuals with ASD. Recent research suggests that children with autism exhibit certain positive social behaviors when interacting with robots compared to their peers that do not interact with robots [2][3][4][5][6]. These positive behaviors include showing emotional facial expressions (e.g., smiling), gesture imitation, and eye gaze attention. Studies show that these behaviors are rare

in children with autism but evidence suggests that robots trigger children to demonstrate such behaviors. These investigations propose that interaction with robots may be a promising approach for rehabilitation of children with ASD. There are several research groups that investigated the response of children with autism to both humanoid robots and non-humanoid toy-like robots in the hope that these systems will be useful for understanding affective, communicative, and social differences seen in individuals with ASD (see Diehl et al., [6]), and to utilize robotic systems to develop novel interventions and enhance existing treatments for children with ASD [13] [14] [15]. Mazzei et al. [16], for example, designed the robot "FACE" to realistically show the details of human facial expressions. A combination of hardware, wearable devices, and software algorithms measured subject's affective states (e.g., eye gaze attention, facial expressions,vital signals, skin temperature and EDA signals), were used for controlling the robot reactions and responses. Reviewing the literature in SAR [5] [6] shows that there are surprisingly very few studies that used an autonomous robot to model, teach or practice the social skills of individuals with autism. Amongst, teaching how to regulate eye-gaze attention, perceiving and expressing emotional facial expressions are the most important ones. Designing robust interactive games and employing a reliable social robot that can sense users' socioemotional behaviors and can respond to emotions through facial expressions or speech is an interesting area of research. In addition, the therapeutic applications of social robots impose conditions on the robot's requirements, feedback model and user interface. In other words, the robot that aims for autism therapy may not be directly used for depression treatment and hence every SAR application requires its own attention, research, and develop-

ment Only a few adaptive robot-based interaction settings have been designed and employed for communication with children with ASD. Proximity-based closed-loop robotic interaction [29], haptic interaction [30], and adaptive game interactions based on affective cues inferred from physiological signals [31] are some of these studies. Although all of these studies were conducted to analyze the functionality of robots for socially interacting with individuals with ASD, these paradigms were limitedly explored and focused on their core deficits (i.e., Facial expression, eye gaze and joint attention skills). Bekele and colleagues [32] studied the development and application of a humanoid robotic system capable of intelligently administering joint attention prompts and adaptively responding based on within system measurements of gaze and attention. They found out that preschool children with ASD have more frequent eye contact toward the humanoid robot agent, and also more accurate respond in joint attention stimulations. This suggests that robotic systems have the enhancements for successfully improve the coordinated attention in kids with ASD. Considering the existing SAR system and the major social deficits that individuals with autism may have, we have designed and conducted robot-based therapeutic sessions that are focused on different aspects of social skills of children with autism. In this thesis we employed NAO which can be remotely controlled to communicate with the children. We conducted two different protocols to examine the social skills of children with autism and provide feedbacks to improve their behavioral responses. The contribution of our work has been introduced in Section 1.4 and the details of the game setting, experiments, modeling and analysis are provided in Chapter 4.

1.3 Music Therapy for ASD

1.4 Thesis Contributions

1.5 Orgizaition

1.5.1 Types of Biometrics

In recent years biometrics moved from simple fingerprinting to many different methods that use various physical and behavioral measures. The characteristics used in each category are as follows:

- Physiological
  - Iris
  - Fingerprint (including nail)
  - Hand (including knuckle, palm, vascular)
  - Face
  - Retina
  - DNA
  - Vein
  - Ear
  - Even Odor, Sweat pore, Lips
- Behavioral

  - Signature
  - Keystroke
  - Voice
  - Gait

- **Identification** is a closed-universe (one-to-many) comparing process for a biometric sample from a given probe against all the known biometric reference templates in the database. In other words, this is the answer to the question "Who am I?" If the acquired sample matches a stored template within an acceptable margin of error, then the identity of the probe is matched to that of the previously stored reference. During the matching process, a set of similarity matching scores are obtained for the probe sample (i.e., one-to-many comparison process). These similarity scores are numerically ranked such that the highest similarity score is first and the smallest similarity score is ranked n, where n is the number of the subjects enrolled in the database. In an ideal case, the highest similarity score is the comparison of the claimed person's biometric with the same person's biometric that was previously stored in the database. The percentage of time that the highest similarity score is the correct match for all individuals, is referred to as the identification rate.

  In order to evaluate the performance of identification, the percentage of time when one of the top-r matches is correct is considered and called as "Cumulative Match score". In other words, the "Cumulative Match Score" curve is the percentage of correct identification versus the rank r. Usually the percentage

of the correct identification for the rank-one is reported as the performance of a biometric system.

- **Verification** is an open-universe (one-to-one) process of comparing a submitted biometric sample against single biometric reference of a single enrollee whose identity or role is being claimed. In other words, this is the answer to the question, "Am I who I claim I am?" The result of the verification is to confirm that the identity is matched or not matched. During the process of matching, a similarity score is computed by the biometric matcher; if the similarity score is higher than a preset threshold T, then the submitted biometric sample is approved to be the same as the biometric reference claimed. If the similarity match score is less than the preset threshold T, then the claimed identity for the submitted biometric is rejected.

  In order to evaluate the verification performance, two kinds of errors can be made by the system: False Match (FM) and False Non-Match (FNM). FM is the error made by deciding that a (claimed) identity is a legitimate one while in reality it is an imposter and FNM is the error made by deciding that a (claimed) identity is not a legitimate while in reality the person is genuine. The frequency rate at which FM occurs is called False Match Rate (FMR), and the frequency rate at which FNM occurs is called False Non-Match Rate (FNMR). The error rates can be evaluated for any threshold T. Therefore, the functions $FMR(T)$ and $FNMR(T)$ give the error rates when the match decision is made at threshold T. The error rates can be plotted against each other as a two-

dimensional curve, (FMR(T), FNMR(T)).

This two-dimensional curve is called Receiver Operating Characteristic (ROC) curve. The ROC curve precisely defines the complete specification of a biometric matcher and shows the trade-off between the FMR and FNMR errors over a wide range of threshold. The biometric matcher can operate using any threshold T which defines a point on the ROC curve. In addition, the ROC can be used to compare the performance of two biometric matchers against each other.

1.5.2 Biometric Market

The research service from the Auto ID & Security business and financial services group highlights growth sectors of notable interest and also provides a comprehensive financial analysis of the biometrics market. The spotlight on security has intensified considerably in the wake of global terror attacks and increasing threats to safety, driving governments across the world to tighten security measures. The demand for sophisticated security solutions is greater now than ever before. Figure 1.1 shows the annual biometric industry revenues for the years 2007-2012 in $m US. As the Figure shows, the annual revenues in the biometric market are growing up with a rate of more than 15% every year. This is due to the huge demand for the applications of biometric technology in different fields.

Figure 1.2 shows the percentage share of the different biometrics in the market in 2006. As the Figure shows, after Fingerprint (43.6%), great attention is paid to face recognition (19.0%) in the biometric market. Advanced face recognition biometrics

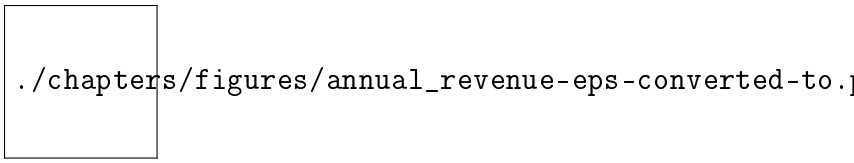

Figure 1.1: *Annual biometric industry revenues for the years 2007-20012.*

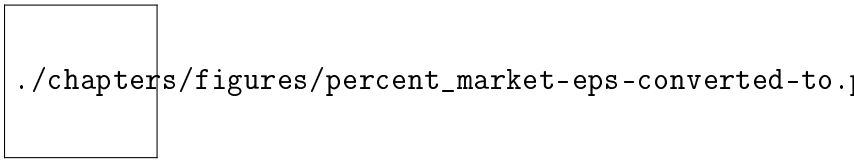

Figure 1.2: *The percent of biometric market by technology in 2006.*

are ideally positioned to address the demand for security solutions and are set to witness a compound annual growth rate (CAGR) of 27.5 percent from $186 million in 2005 to $1021.1 million in 2012 [**?**].

Enhanced credibility of this technology combined with its rapidly growing awareness is also likely to provide a strong impetus to growth of the face recognition biometrics market throughout the forecast period. Concrete evidence in the form of successful deployments has also helped contribute to continued market growth.

Human Computer Interaction

Human-computer interaction (HCI) is the study of interaction between people (users) and computers. To achieve efficient and user-friendly interaction, the human body part (e.g., the face) could be considered as a natural input device. For instance the movements of the face can be used in human tracking system. We recently developed an efficient tracking system of people based on their facial skin and body (cloth)

Category Area	Applications
Face ID	Voter registration, Driver licenses, national ID, immigration
Access Control	Building/room access, computer access
Security	Terrorist alert, secure flight boarding system
Surveillance	Advanced video surveillance, nuclear plants surveillance, neighborhood watch, power grid surveillance, portal control
Smart	Cards stored valued security, user authentication
Law Enforcement	Crime stopping and suspect alert, shoplifter recognition, suspect background check, post event analysis
Face-based database	Face-based search and retrieval
Multimedia management	Indexing, segmentation, classification, or event detection
Human computer interaction	Interactive gaming, animation

Table 1.1: *Typical applications of face.*

colors using a single video camera [**?**]. Also, the tracked faces can be used as first step to localize the location of faces, in video images, for face recognition. Facial expression recognition is the ability of computers to understand human emotions. Cohen *et al.* [**?**] reported on several Advances they have made in building a system for classifying facial expression from continuous video input. They used Bayesian network classifiers for classifying expressions from video. Another application of HCI is realistic synthesis and animation of faces which are widely used in the video and motion picture industries as well as the video game industry. Hong *et al.* [**?**] designed a system that provides functionalities for 3-D face modeling and animation with the help of user interactions. Text and speech streams can be used to drive the face animation which is used in computer aided education.

- Intrinsic variations are independent of any external sources and are due to the physical nature of the human face.

- Extrinsic variations are caused by the sources that do not depend on the human

Variation in appearance	**Source**	**Effect/possible task**
Extrinsic	Viewing geometry, Illumination, Imaging process Other objects	Head Pose light variations, shadow, self shadow Resolution, scale, focus, sampling Occlusion, shadowing, indirect illumination, hair, make-up, surgery
Intrinsic	Identity, Facial expression, Age, Sex, Speech	Identification, known-unknown Inference of emotion or intension Estimating age Decide if male or female Lip reading

Table 1.2: *Variations in facial appearance Inter-person and intra-person variations.*

face or any subject under test and are due to the factors such illumination, viewing geometry, and the imaging process.

Table 1.2 summarizes these two types of variations and their effects on face recognition.

Among these effects, illumination, variations in pose, aging, and facial expressions are the most challenging for face recognition.

- **illumination:** Changes in lighting conditions, e.g., indoor or outdoor, under which the facial images are captured, affect the accuracy of face recognition. Variations in illumination could be caused either by variations in the light source or by variations in physical parameters of the cameras and the capturing devices. A solution for this problem is by utilizing the 3-D surface information of the face. So, by having the 3-D model of the face surface, the problem reduces to matching the surface geometry of two faces which are invariant under the effect of illumination.

- **Head Pose:** Pose variation is another challenging problem in face recognition. The variations in pose could be because of the changes in viewing angle of the camera which causes pose variation in the 2-D or 3-D captured face image. Because face is a 3-D object, 2-D face recognition under the effect of pose variations is difficult, while having the 3-D face data, the problem of pose variation can be handled either in 3-D versus 3-D face recognition or 2-D versus 3-D.

- **Facial Expressions:** The development of robust face recognition algorithm insensitive to facial expression is one of the biggest challenges of current research in this field. The change in the face appearance due to its non-rigid structure makes modeling and analyzing the facial expressions difficult. In addition, facial expressions vary from person to person, which makes the task of modeling the facial expressions more difficult.

- **Aging Effect:** Aging is the inherent problem of face recognition because face is an identifier that changes with age and the aging effect cannot be controlled or ignored. The facial aging effects are manifested in different forms in different ages. It is manifested as changes in the shape of the cranium from infancy to teenage while during the adulthood it is demonstrated as changes in the skin texture. Thus, because facial aging has different sources, having a unified solution for this problem is difficult.

Another challenge for face recognition is the need for an evaluation standard for measuring recognition performance under different environments and conditions. As a result of this necessity, an independent government evaluation standard was born,

which is called, Face Recognition Vendor Tests (FRVT). FRVT was developed to provide evaluations of commercially available and prototype face recognition technologies. These evaluations are designed to provide U.S. government and law enforcement agencies with information to assist them in determining where and how facial recognition technology can best be deployed. In addition, FRVT results help identify future research directions for the face recognition community. In the past, many factors have been evaluated in FRVT 2002 [**?**]. For example, in a verification test with reasonably controlled lighting, when the gallery consisted of 37,437 individuals with one image per person and the probe set of 74,854 probes with two images per person, the best three systems averaged a verification rate of 90% at false accept rate of 1%, 80% at false accept rate of 0.1%, and 70% at false accept rate of 0.01%. This level of accuracy may be suitable for access control with a small database of hundreds of people but not for a security system at airports where the number of passengers is much larger. When evaluating the performance with respect to pose changes with a database of 87 individuals, the best system achieved an identification rate of 42% for faces within ± 45 degrees of panning and 53% within ± 45 degrees of tilting. Lighting changes between outdoor probe images and indoor gallery images degrade the best systems from a verification rate of 90% to 60% at a false accept rate of 1%.

1.6 Proposed Face Modeling and Recognition System

In Face Recognition Grand Challenge (FRGC) contest, three contenders for improving face recognition algorithms were considered: high resolution images, three-dimensional (3-D) face recognition, and multiple still images. With the 3-D data, the two main challenges of face recognition, pose variation and illumination, can be handled easily. This is due to the fact that the 3-D shape of a person's face is not affected by changes in head orientation and lighting. Hence, 3-D face recognition has the potential of improving the recognition performance under these conditions [**?**]. Nevertheless, a pure 3-D face recognition system has its own following limitations.

- Capturing the 3-D face data either by a range scanner or by a stereo-based system is slow and expensive with the current technology.
- Capturing such a 3-D data is intrusive.
- Extraction of facial landmarks in 3-D is a very challenging task.
- Shape matching techniques are complex and time consuming.
- Lack of the texture cue in captured 3-D range data.

3-D Face Recognition Using 3-D Ridge Images

In this dissertation, we present a novel method for 3-D face recognition (shape matching) based on ridge lines extracted from the 3-D range facial images. Compared to

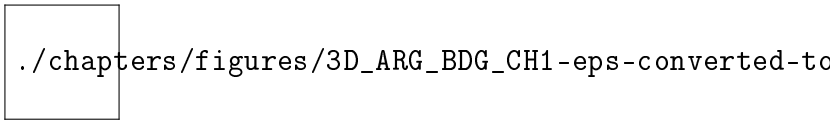

Figure 1.3: *The general block diagram of our system for multi-modal face recognition based on 3-D ARG models.*

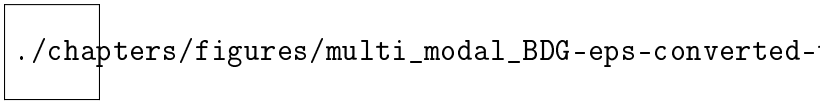

Figure 1.4: *The general block diagram of our system for multi-modal face recognition based on ridge images and 2-D ARG models.*

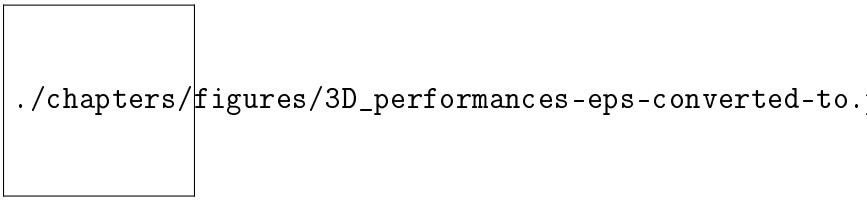

Figure 1.5: *Comparison between the three categories of algorithms for 3-D face recognition, performance of the system versus complexity.*

Figure 1.6: *Samples of extracted ridge images.*

other shape matching based approaches for 3-D face recognition, such as [? ? ? ?], our approach is faster and requires less computations. This reduction in computations is due to the fact that we only use the points around the important facial regions on the face (i.e., the eyes, the nose, and the mouth) and ignore other surface patches on the face during the matching process. These points correspond to the extreme ridge points on the considered surface. An extreme ridge point is a point where the principal curvature k_{max}, has large positive value. There are different approaches to locate the ridges, here we threshold the k_{max} values to find these points. Figure 1.6 shows few examples of the ridge images obtained by thresholding the k_{max} values. These are 3-D binary images that show the location of the ridge lines on the surface of the face. In this work, the number of the points in a ridge image of the face is $12\% \pm 2\%$ of the total number of points that cover the face. For matching the ridge images (probe image versus gallery image), either the Hausdorrf Distance or the ICP method can be used.

1.7 Dissertation Contributions

The major contributions of this dissertation are as follows:

- Improving the Active Shape Model for 2-D facial features extraction from color

image. We present solutions for some of the limitations of Active Shape Model (ASM) to extract facial feature extraction in color images.

- Developing an algorithm for 3-D facial feature extraction from range data. Extracting 3-D facial features from 3-D range images is more difficult compared to 2-D facial feature extraction, because of the lack of texture in range images. In this dissertation, we develop an algorithm for extracting three facial feature points (i.e., the inner corners of the two eyes and the tip of the nose) from facial range images. These points are used to initially align the ridge images during the matching process.

- Developing an algorithm for 3-D face modeling and recognition based on ridge images. The ridge lines in the range image carry the most important distinguishing information of the 3-D face and have high potential for face recognition. We develop a system for 3-D face recognition based on ridge lines. For matching the ridge images of two faces (probe and gallery), the Hausdorff and Iterative Closest Points are utilized.

- Developing a novel algorithm for 3-D face recognition based on Attributed Relational Graphs (ARG). The nodes of the graph represent the facial landmark points. A set of attributes are extracted using Gabor filters and assigned to each node of the graph. Also, a set of features that defines the mutual relations between the edges of the graph are extracted and used to increase the performance of the graph model for face recognition.

- Developing a multi-modal technique based on the Dempster-Shafer theory of evidence and the weighted sum rule for fusion at the score level.

1.8 Dissertation Outline

This thesis is organized as follows: In Chapter two, we present related work for facial features extraction, two dimensional (2-D), three dimensional (3-D), and multi-modal (2-D + 3-D) face recognition. Chapter three explains our algorithm for 2-D facial feature extraction from frontal face images (i.e., Improved ASM) and our algorithm for 3-D facial feature extraction (i.e., the extraction of the three feature points) along with the experimental results. Chapter four presents our approach for 3-D face modeling and recognition based on ridge images. Chapter five describes our multi-modal face modeling and recognition (2-D/3-D) based on attributed relational graphs along with the experiments. In addition, we present two fusion techniques for combining the 2-D and 3-D modalities in this chapter. Finally in Chapter six, we present the conclusion and the future research directions.

Chapter 2

Related Works

2.1 Autism

Individuals with autism spectrum disorder experience verbal and nonverbal communication impairments, including motor control, emotional facial expressions, and eye gaze attention. Oftentimes, individuals with high-functioning autism have deficits in different areas, such as (1) language delay, (2) difficulty in having empathy with their peer and understanding others emotions (i.e. facial expressions recognition.), and more remarkably (3) joint attention (i.e. eye contact and eye gaze attention). Autism is a disorder that appears in infancy [23]. Although there is no single accepted intervention, treatment, or known cure for ASDs, these individual will have more successful treatment if ASD is diagnosed in early stages. At the first glance at the individual with autism, you may not notice anything odd, however after trying to talk to her/him, you will understand something is definitely not right [77]. S/He may not make eye contact with you and avoid your gaze and fidget, rock her/his body and

bang her/his head against the wall [77]. In early 1990s, researchers in the University of California at San Diego aimed to find out the connections between autism and nervous system (i.e. mirror neurons). Mirror neuron [77] is a neuron that is activated either when a human acts an action or observes the same action performed by others. As these neurons are involved with the abilities such as empathy and perception of other individual's intentions or emotions, they came up with malfunctioning of mirror neuron in individuals with ASD [77]. There are several studies that focus on the neurological deficits of individuals with autism and studying on their brain activities. Figure 2-1 demonstrates the areas in the brain that causes the reduce mirror neuron activities in individuals with autism. Individuals with autism might also have several other unusual social developmental behaviors that may appear in infancy or childhood. For instance children with autism show less attention to social stimuli (e.g. facial expressions, joint attention), and respond less when calling their names. Compared with typically developing children, older children or adults with autism can read facial expressions less effectively and recognize emotions behind specific facial expressions or the tone of voice with difficulties [26]. In contrast to TD individuals, children with autism (i.e. high-functioning, Asperger syndrome) may be overwhelmed with social signals such as facial behaviors and expression and complexity of them and they suffer from interacting with other individuals, therefore they would prefer to be alone. That is why it would be difficult for individuals with autism to maintain social interaction with others [28]. In order to diagnose and asses the aspects and score the social skill level of an individual with autism, several protocols are available. One of the commercially available protocols is called Autism Diagnostic Observation

Schedule (ADOS) [65] that consists of four modules and several structured tasks that are used to measure the social interaction levels of the subject and examiner. We are inspired by ADOS in designing our intervention protocols later described in Chapter 4. Hence, we briefly review ADOS in the next section.

2.1.1 Turn-Taking

Turn-taking is a type of organization in conversation and discourse where participants speak one at a time in alternating turns. In practice, it involves processes for constructing contributions, responding to previous comments, and transitioning to a different speaker, using a variety of linguistic and non-linguistic cues.[1]

While the structure is generally universal,[2] that is, overlapping talk is generally avoided and silence between turns is minimized, turn-taking conventions vary by culture and community.[3] Conventions vary in many ways, such as how turns are distributed, how transitions are signaled, or how long is the average gap between turns.

In many contexts, conversation turns are a valuable means to participate in social life and have been subject to competition.[4] It is often thought that turn-taking strategies differ by gender; consequently, turn-taking has been a topic of intense examination in gender studies. While early studies supported gendered stereotypes, such as men interrupting more than women and women talking more than men,[5] recent research has found mixed evidence of gender-specific conversational strategies, and few overarching patterns have emerged.[6]

https://en.wikipedia.org/wiki/Turn-taking

2.1.2 Emotion EDA

Emotion is an intense mental experience often manifested by rapid heartbeat, breathing, sweating, and facial expressions. Emotion recognition from these physiological signals is a challenging problem with interesting applications such as developing wearable assistive devices and smart human-computer interfaces. This paper presents an automated method for emotion classification in children using electrodermal activity (EDA) signals. The time-frequency analysis of the acquired raw EDAs provides a feature space based on which different emotions can be recognized. To this end, the complex Morlet (C-Morlet) wavelet function is applied on the recorded EDA signals. The dataset used in this paper includes a set of multimodal recordings of social and communicative behavior as well as EDA recordings of 100 children younger than 30 months old. The dataset is annotated by two experts to extract the time sequence corresponding to three main emotions including "Joy", "Boredom", and "Acceptance". The annotation process is performed considering the synchronicity between the children's facial expressions and the EDA time sequences. Various experiments are conducted on the annotated EDA signals to classify emotions using a support vector machine (SVM) classifier. The quantitative results show that the emotion classification performance remarkably improves compared to other methods when the proposed wavelet-based features are used. https://www.sciencedirect.com/science/article/pii/S0957417418303609

2.1.3 Motor control

Motor control is the systematic regulation of movement in organisms that possess a nervous system. Motor control includes movement functions which can be attributed to reflex,[1]. Motor control as a field of study is primarily a sub-discipline of psychology or neurology.

Recent psychological theories of motor control present it as a process by which humans and animals use their brain/cognition to activate and coordinate the muscles and limbs involved in the performance of a motor skill. From this mixed psychological perspective, motor control is fundamentally the integration of sensory information, both about the world and the current state of the body, to determine the appropriate set of muscle forces and joint activations to generate some desired movement or action. This process requires cooperative interaction between the central nervous system and the musculoskeletal system, and is thus a problem of information processing, coordination, mechanics, physics, and cognition.[2][3] Successful motor control is crucial to interacting with the world, not only determining action capabilities, but regulating balance and stability as well.

The organization and production of movement is a complex problem, so the study of motor control has been approached from a wide range of disciplines, including psychology, cognitive science, biomechanics and neuroscience. While the modern study of motor control is an increasingly interdisciplinary field, research questions have historically been defined as either physiological or psychological, depending on whether the focus is on physical and biological properties, or organi-

zational and structural rules.[4] Areas of study related to motor control are motor coordination, motor learning, signal processing, and perceptual control theory. https://en.wikipedia.org/wiki/Motor_control

2.2 ADOS

The Autism Diagnostic Observation Schedule (ADOS) is a standardized protocol for observing the social and communicative behaviors associated with autism. Eight tasks have contained in ADOS, as shown in the table below. 20-30 minutes are required for an examiner to complete the entire tasks [65]. As shown in TABLE 2-1, each task contains one or few aspects of social skills including turn taking (refers to the process by which people in a conversation decide who is to speak next), joint attention, reading emotions etc. Right after the interview, examiner would provide a general ratings based on the observation in all the tasks which have been targeted to code. ADOS contains four modules that are designed for specific age range and set of social developmental abilities. Examiner may use 'Module 1' if the child uses a little or no phrase speech however if s/he utilizes phrase speech but do not speak fluently 'Module 2' may be employed. Some examples of Modules 1 or 2 are responding to name, social smile, and bubble play. 'Module 3' is used for younger children who are verbally fluent and 'Module 4' is employed for adolescents and adults with fluent verbal skills. Modules 3 or 4 can include communication, and exhibition of empathy or comments on others' emotions. Considering these four modules, ADOS can provide scores regarding these four areas (1) Reciprocal social interaction, (2) Communication/language, (3)

Stereotyped/restricted behaviors, and (4) Mood and non-specific abnormal behaviors. In our study we employed ADOS and some tasks described in it for introducing new robot-based games and social interaction to children that will be explained in Chapter 4 section 2.3.

Chapter 3

Xylo-Bot: A Interactive Human-Robot Music Teaching System Design

A novelty Interactive human-robot music teaching system design is presented in this chapter. In order to make robot play xylophone properly, several things need to be done before that. First is to find a proper xylophone with correct timber; second, we have to make the xylophone in a proper position in front of the robot that makes it to be seen properly and be reached to play; finally, design the intelligent music system for NAO.

3.1 NAO: A Humanoid Robot

We used a humanoid robot called NAO developed by Aldebaran Robotics in France [ref]. NAO is 58 cm (23 inches) tall, with 25 degrees of freedom this robot can conduct most of the human behaviors. It also features an onboard multimedia system including, four microphones for voice recognition, and sound localization, two speakers for text-to-speech synthesis, and two HD cameras with maximum image resolution 1280 x 960 for online observation. As shown in Figure 4-1, these utilities are located in the middle of the forehead and the mouth area. NAO's computer vision module includes facial and shape recognition units. By using Choregraphe software (Shown in Figure 4-2), researcher can easily control NAO remotely. Inside the user interface we have access to NAO's cameras. It is also easy to control different joints of the robot (see Figure 4-3). This allows the operator to control and monitor the different activities of robots online. The robot arms have a length of approximately 31 cm. Each arm have five degrees of freedom and is equipped with the sensors to measure the position of each joint. To determine the pose of the instrument and the beaters' heads the robot analyzes images from the lower monocular camera located in its head, which has a diagonal field of view of 73 degree. These dimensions allows us to choose a proper instrument presented in next section.

3.2 Accessories

Due to the size limitation of the toy xylophone, we have to design some accessories for robot to able play.

3.2.1 Xylophone: A Toy for Music Beginner

Attach the picture of xylophone and describe the frequency of all notes. We choose a Sonor Toy Sound SM soprano-xylophone with 11 sound bars of 2 cm in width. The instrument has a size of xx cm x xx cm x xx cm, including the resonateing body. The smallest sound bar is playable in an area of 2.8 cm x 2 cm, the largest in an area of 4.8 cm x 2 cm. The instrument is diatonically tuned in C-Major/a-minor. The beaters/mallet, we use the pair which come with the xylophone with a modified 3D printed grips (details in next subsection) to allow the robot's hands to hold them properly. The mallets are approximately 21 cm in length include a head of 0.8 cm radius.

3.2.2 Mallet Gripper Design

3D printed, need to measure some numbers and list them here, attach the SolidWorks screen shot and actual pictures.

3.2.3 Instrument Stand Design

Laser cut, made of wood, need measurement of all dimensions, attach the SolidWorks screen shot and actual pictures.

3.3 Music Teaching System Design

After all the preparation, we start to design the system, including joint trajectory, vision control and audio feedback

3.3.1 Joint Trajectory

Calibration of kinematic parameters. Try to explain it better at some point, if possible describer the future work may implemented using vision feedback system.

3.3.2 Auditory Feedback System

The purpose of this system is to provide a back and forth interaction using music therapy to teach kid social skills and music knowledge.

Short Time Fourier Transform

The short-time Fourier transform (STFT), is a Fourier-related transform used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time.[1] In practice, the procedure for computing STFTs is to divide a longer time signal into shorter segments of equal length and then compute the Fourier transform separately on each shorter segment. This reveals the Fourier spectrum on each shorter segment. One then usually plots the changing spectra as a function of time. Use wiki to attach more pics and more info here. https://en.wikipedia.org/wiki/Short-time$_{F}ourier_{t}ransform$

3.3.3 Dialog System

Speech Recognition

http://doc.aldebaran.com/2-1/naoqi/audio/alspeechrecognition.html

reason to design the dynamic feedback, NPL may want to have it here.

3.4 Summary

Chapter 4

X-Elophone: A Revolution of Xylophone

4.1 More Sound, More Possibilities

reason why need this design. Due to the limitation of keys. This provides more possibility for different timber and major minor keys. That allows this system to play more customized song which kids love.

4.1.1 Components Selection

Piezo Vibration Sensor

The LDT0-028K is a flexible component comprising a 28 m thick piezoelectric PVDF polymer film with screen-printed Ag-ink electrodes, laminated to a 0.125 mm polyester substrate, and fitted with two crimped contacts. As the piezo film is displaced from

the mechanical neutral axis, bending creates very high strain within the piezopolymer and therefore high voltages are generated. When the assembly is deflected by direct contact, the device acts as a flexible "switch", and the generated output is sufficient to trigger MOSFET or CMOS stages directly. If the assembly is supported by its contacts and left to vibrate "in free space" (with the inertia of the clamped/free beam creating bending stress), the device will behave as an accelerometer or vibration sensor. Adding mass, or altering the free length of the element by clamping, can change the resonant frequency and sensitivity of the sensor to suit specific applications. Multi-axis response can be achieved by positioning the mass off center. The LDTM-028K is a vibration sensor where the sensing element comprises a cantilever beam loaded by an additional mass to offer high sensitivity at low frequencies. https://cdn.sparkfun.com/datasheets/Sensors/ForceFlex/LDT$_S$*eries.pdf*

Also have to show the circuit, how to design this and attach the figure from here https://www.sparkfun.com/datasheets/Sensors/Flex/MSI-techman.pdf page 39

Op-Amp

An operational amplifier (often op-amp or opamp) is a DC-coupled high-gain electronic voltage amplifier with a differential input and, usually, a single-ended output.[1] In this configuration, an op-amp produces an output potential (relative to circuit ground) that is typically hundreds of thousands of times larger than the potential difference between its input terminals. Operational amplifiers had their origins in analog computers, where they were used to perform mathematical operations in many linear, non-linear, and frequency-dependent circuits. The popularity of the op-amp as

a building block in analog circuits is due to its versatility. By using negative feedback, the characteristics of an op-amp circuit, its gain, input and output impedance, bandwidth etc. are determined by external components and have little dependence on temperature coefficients or engineering tolerance in the op-amp itself. Op-amps are among the most widely used electronic devices today, being used in a vast array of consumer, industrial, and scientific devices. Many standard IC op-amps cost only a few cents in moderate production volume; however, some integrated or hybrid operational amplifiers with special performance specifications may cost over US 100 in small quantities.[2] Op-amps may be packaged as components or used as elements of more complex integrated circuits. The op-amp is one type of differential amplifier. Other types of differential amplifier include the fully differential amplifier (similar to the op-amp, but with two outputs), the instrumentation amplifier (usually built from three op-amps), the isolation amplifier (similar to the instrumentation amplifier, but with tolerance to common-mode voltages that would destroy an ordinary op-amp), and negative-feedback amplifier (usually built from one or more op-amps and a resistive feedback network). https://en.wikipedia.org/wiki/Operational$_amplifierhttps$: $//ww1.microchip.com/downloads/en/DeviceDoc/21733j.pdf$

Multiplexer

In electronics, a multiplexer (or mux) is a device that selects between several analog or digital input signals and forwards it to a single output line.[1] A multiplexer of $2^n 2^n inputshasnnselectlines, whichareusedtoselectwhichinputlinetosendtotheoutput.[2]Multiplexe$

$input, single-outputswitch, andademultiplexerasasingle-input, multiple-outputswitch.[3]Thesc$

$to-1multiplexeronthelefttandanequivalentswitchontheright.Theselselwireconnectsthedesiredinp$

$poleoctal-throwanalogswitch(SP8T)suitableforuseinanalogordigital8:1multiplexer/demultiple$

$//en.wikipedia.org/wiki/Multiplexerhttps://cdn.sparkfun.com/assets/learn_tutorials/5/5/3/7$

Arduino UNO

The Arduino Uno is an open-source microcontroller board based on the Microchip ATmega328P microcontroller and developed by Arduino.cc.[2][3] The board is equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits.[1] The board has 14 Digital pins, 6 Analog pins, and programmable with the Arduino IDE (Integrated Development Environment) via a type B USB cable.[4] It can be powered by the USB cable or by an external 9-volt battery, though it accepts voltages between 7 and 20 volts. It is also similar to the Arduino Nano and Leonardo.[5][6] The hardware reference design is distributed under a Creative Commons Attribution Share-Alike 2.5 license and is available on the Arduino website. Layout and production files for some versions of the hardware are also available.

The word "uno" means "one" in Italian and was chosen to mark the initial release of the Arduino Software.[1] The Uno board is the first in a series of USB-based Arduino boards,[3] and it and version 1.0 of the Arduino IDE were the reference versions of Arduino, now evolved to newer releases.[4] The ATmega328 on the board comes preprogrammed with a bootloader that allows uploading new code to it without the use of an external hardware programmer.[3]

While the Uno communicates using the original STK500 protocol,[1] it differs from

all preceding boards in that it does not use the FTDI USB-to-serial driver chip. Instead, it uses the Atmega16U2 (Atmega8U2 up to version R2) programmed as a USB-to-serial converter.[7] https://en.wikipedia.org/wiki/Arduino$_{U}noshowblockdiagramofthecode$:

4.1.2 ChucK: A On-the-fly Audio Programming Language

https://www.researchgate.net/profile/Ge$_{W}ang9/publication/259326122_{T}he_{C}hucK_{P}rogramming_{L}a$

$Timed_{O}n-the-fly_{E}nvironmentality/links/0c96052b02acb79c2c000000.pdfbrieflydescribethelan$

$butonlytotheextentthatweareabletoexpresstothecomputerwhattodo, andhowtodoit.Tothisend, thep$

$specific]languagescanbringadditionalexpressiveness, conciseness, andperhapsevendifferentway$

$purposeprogramminglanguagetailoredforcomputermusic.Thegoalistocreatealanguagethatisexpre$

$basedconcurrentprogrammingmodelthatallowsprogrammerstoflexiblyandpreciselycontrolthefl$

$timed]), andfacilitiestodevelopprogramson-the-fly-astheyrun.AChucKianapproachtolivecodi$

$the-flyprogramming, tovisualizeandmonitorChucKprogramsinreal-time, andtoprovideaplatf$

$mediatedensembles.Additionalapplicationsarealsodescribed, includingclassrooms, livecodingare$

$basedprogrammingmechanism(bothlanguageandunderlyingimplementation)forultrapreciseaudi$

$timeaudioanalysis.2)Anonpreemptive, time/event-basedconcurrentprogrammingmodelthatprov$

$the-fly.Thisrapidprototypingmentalityhaspotentiallywideramificationsinthewaywethinkabout$

$timeaudio), aswellasnewparadigmsandpracticesincomputer-mediatedliveperformance.4)TheAu$

$andhowthesetwodisciplinescanreinforceeachother.$

VITA

Mohammad Hossein Mahoor was born in Estahban, Fars Province, Iran, on January 27, 1975. He received his elementary education at Shahid Faghihi Elementary School, his secondary education at Dr. Shariati Middle School, and his high school education at Shahid Beheshti High School. In September 1992, he was admitted to the University of Petroleum Industry (Former Abadan Institute of Technology, A.I.T.) in Ahwaz, Iran, from which he was graduated with the B.S. degree in Electrical Engineering with first-class honor in September 1995. He continued his graduate studies in Sharif University of Technology and was awarded M.S. degree in Biomedical Engineering in October 1998.

In August 2003, he was admitted to the Graduate School of the University of Miami, where he was granted the degree of Doctor of Philosophy in Electrical and Computer Engineering in December 2007.

Permanent Address: 1251 Memorial Dr. #EB406, Coral Gables, Florida 33124