Summer 2016

Computational Modeling of Facial Response for Detecting Differential Traits in Autism Spectrum Disorders

Manar D. Samad
Old Dominion University, manar.eee@gmail.com

COMPUTATIONAL MODELING OF FACIAL RESPONSE FOR DETECTING DIFFERENTIAL TRAITS IN AUTISM SPECTRUM DISORDERS

by

Manar D. Samad
B.S. June 2007, Bangladesh University of Engineering & Technology, Bangladesh
M.S. January 2011, The University of Calgary, Canada

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ELECTRICAL AND COMPUTER ENGINEERING

OLD DOMINION UNIVERSITY
August 2016

Approved by:

Khan M. Iftekharuddin (Director)

Jiang Li (Member)

Lee A. Belfore (Member)

Michel A. Audette (Member)

John W. Harrington (Member)

ABSTRACT

COMPUTATIONAL MODELING OF FACIAL RESPONSE FOR DETECTING DIFFERENTIAL TRAITS IN AUTISM SPECTRUM DISORDERS

Manar D. Samad
Old Dominion University, 2016
Director: Dr. Khan M. Iftekharuddin

This dissertation proposes novel computational modeling and computer vision methods for the analysis and discovery of differential traits in subjects with Autism Spectrum Disorders (ASD) using video and three-dimensional (3D) images of face and facial expressions. ASD is a neurodevelopmental disorder that impairs an individual's nonverbal communication skills. This work studies ASD from the pathophysiology of facial expressions which may manifest atypical responses in the face. State-of-the-art psychophysical studies mostly employ naïve human raters to visually score atypical facial responses of individuals with ASD, which may be subjective, tedious, and error prone. A few quantitative studies use intrusive sensors on the face of the subjects with ASD, which in turn, may inhibit or bias the natural facial responses of these subjects. This dissertation proposes non-intrusive computer vision methods to alleviate these limitations in the investigation for differential traits from the spontaneous facial responses of individuals with ASD. Two IRB-approved psychophysical studies are performed involving two groups of age-matched subjects: one for subjects diagnosed with ASD and the other for subjects who are typically-developing (TD). The facial responses of the subjects are computed from their facial images using the proposed computational models and then statistically analyzed to infer about the differential traits for the group with ASD. A novel computational model is proposed to represent the large volume of 3D facial data in a small pose-invariant Frenet frame-based feature space. The inherent pose-invariant property of the proposed features alleviates the need for an expensive 3D face registration in the pre-processing step. The proposed modeling framework is not only computationally efficient but also offers competitive performance in 3D face and facial expression recognition tasks when compared with that of the state-of-the-art methods. This computational model is applied in the first experiment to quantify subtle facial muscle response from the geometry of 3D facial data. Results show a statistically significant asymmetry in specific pair of facial muscle activation ($p<0.05$) for the group with ASD, which suggests the presence of a psychophysical trait (also known as an

'oddity') in the facial expressions. For the first time in the ASD literature, the facial action coding system (FACS) is employed to classify the spontaneous facial responses based on facial action units (FAUs). Statistical analyses reveal significantly ($p<0.01$) higher prevalence of smile expression (FAU 12) for the ASD group when compared with the TD group. The high prevalence of smile has co-occurred with significantly averted gaze ($p<0.05$) in the group with ASD, which is indicative of an impaired reciprocal communication. The metric associated with incongruent facial and visual responses suggests a behavioral biomarker for ASD. The second experiment shows a higher prevalence of mouth frown (FAU 15) and significantly lower correlations between the activation of several FAU pairs ($p<0.05$) in the group with ASD when compared with the TD group. The proposed computational modeling in this dissertation offers promising biomarkers, which may aid in early detection of subtle ASD-related traits, and thus enable an effective intervention strategy in the future.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disability that is characterized by behavioral restrictions, cognitive deficits, and impairments in social and interpersonal skills. Autism and Developmental Disabilities Monitoring (ADDM) Network, USA reports that the prevalence of ASD among 8-year-old children is 1 in 68 from 11 states in the surveillance year 2010 [1]. The range of the prevalence is reported to be from 21.9 per 1,000 in New Jersey to 5.7 per 1,000 in Alabama. The report, published on March 28, 2014, reveals a rapid increase of 23% from 2002 to 2006 and a 78% increase from 2002 to 2008 in the prevalence rate of population with ASD. This is significant because ASD is pervasive, and will continue to have lifetime impacts on the health and function of an affected individual. In addition, ASD entails four to six times more medical expenses than that of non-ASD individuals [2] with a substantial amount of treatment and special education services. Since the prevailing rate is increasing rapidly, there is a need for extensive research and development efforts to identify the pathophysiological targets and minimize functional limitations and disabilities. Unfortunately, the cause or origin of the disorder is not well understood, and the relevant research findings are often inconclusive and inconsistent. This is because of the inherent complexity and heterogeneity of the spectrum disorder as it is found co-morbid with multiple clinical conditions such as anxiety, phobia, hyperactivity and so on [3]. The existing research efforts can be categorized into studies such as clinical, psychological, radiological, and behavioral to understand the etiology and pathophysiology of the disorder. However, the complexity of the disorder requires multidisciplinary research efforts that combine multiple areas of expertise in more systematic, thoughtful, and innovative ways. This, in turn, underscores the need to leverage scientific resources, technologies, and innovations to accommodate the needs for next generation diagnosis and treatment measures.

1.1 Diagnosis and Screening for ASD

The onset of ASD usually occurs during the early developmental period of a child with subtle deviances noted after the six-month period [4]. Unfortunately, there is no reliable

biomarker for the diagnosis of ASD [5]. Therefore, this disorder is commonly diagnosed through the direct visual observation of atypical verbal and nonverbal behaviors. While early identification of ASD is highly desirable for the implementation of effective intervention strategies, it can be challenging to visually identify subtle abnormalities displayed by the child. Harrington et al. systematically delineate the aspects of understanding, diagnosing, screening, and treating of children with ASD [3]. A clinician needs to be knowledgeable of the signs and symptoms of ASD following the changes that may occur in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [6]. The DSM-5 suggests multiple criteria of deficits in individuals with ASD [7]. Children who meet the DSM-5 criteria are typically diagnosed by the age of two to three. Many of the impairments related to ASD appear in the nonverbal and joint attention skills of the child, which are screened using the Modified Checklist for Autism in Toddlers (M-CHAT) [8]. Unfortunately, the M-CHAT is not suitable for children over three years of age. On the other hand, there can be subtle cases of ASD, which may take a longer time to diagnose. As the child with ASD gradually attempts to engage in social interactions, these subtle abnormalities can lead to complex behavioral impairments such as lack of social communication skills, deficits in perceiving emotions, and the inability to respond with appropriate facial expressions. The impairments related to ASD may not be evident among children until the school-age involving higher social interactions. Therefore, early diagnosis can be challenging for physicians who rely on observational methods to identify the abnormalities displayed by the child. The diagnosis of ASD during the early childhood period is crucial in order to achieve an effective intervention.

Children diagnosed with ASD are known to develop varying manifestations of symptoms and severity [9, 10]. Therefore, a child diagnosed with ASD needs follow-up clinical screenings to detect and measure atypical traits in behavior, which may emerge over the developmental period. The atypical traits associated with ASD are often termed as broad autism phenotype (BAP), which are targeted to identify and evaluate the deviances in ASD from typical development [11]. Researchers believe that the study on BAP may help finding useful information regarding the etiology and early identification of ASD [12]. For the identification and evaluation of BAP-related traits among pre-school and school-aged children with ASD, there are measurement tools developed by clinicians and researchers. These tools are based on interviews [13], visual observations [14], rating scales and questionnaires [8, 15]. Unfortunately, all these measurement tools are subjective and require qualitative visual assessments. Dawson et al. propose an effective measurement tool known

as the Broader Phenotype Autism Symptom Scale (BPASS). The BPASS is used to rate ASD-related traits affecting four developmental skills of school-aged children: social motivation, expressiveness, conversational skills, and range of interests [14]. The BPASS requires trained clinicians to interact with children with ASD to observe and rate ASD-related traits for four items of expressiveness: eye contact, vocal prosody, facial expressions, and social smile.

1.2 Problem Statement

The above discussions demand both effective computational models and extensive psychophysical studies to propose objective and measurable targets for screenings and interventions of ASD. Several experiments are performed to validate the proposed computational models and investigate for differential traits from facial response data of individuals with ASD. The experiments are proposed to achieve the following goals related to one or more problem statements.

The first goal of this dissertation is to obtain a computational model for generalized representation and analysis of 3D facial data. A generalized computational model for 3D facial data is proposed to measure both the facial shape for identity retrieval and the facial muscle response for expression analysis. However, the large volume of data in 3D faces limits effective extraction of essential features for the computational modeling, which needs to be addressed without compromising the performance of the computational model. Experiments on several benchmarked 3D facial datasets are performed to demonstrate the effectiveness of the generalized computational model.

The second goal of this dissertation is to design and develop experimental protocols and setup in order to overcome several limitations in existing psychophysical studies on subjects with ASD. The existing studies mostly employ naïve human raters to perform retrospective analysis of videotaped facial images of subjects with ASD [16, 17]. The elicited or spontaneous facial responses in subjects with ASD are often found to be attenuated and covert, making them too subtle and subjective to be perceived by human eyes [18]. The intrusive administration of electrophysiological sensors (e.g., EMG) on facial skin may inhibit and bias the natural response of the subjects with ASD. In other studies, the ability to imitate six target prototypical expressions is studied using computer vision-based methods in subjects with ASD [19] and schizophrenia [20]. Unlike spontaneous facial expressions, the task of imitation does not reflect the natural psychophysical response of an individual due to social or emotional trigger. In general, an ergonomic experimental setup

needs to be designed for an effective psychophysical study to ensure minimal invasion and stress while yielding maximum engagement and elicitation from subjects with ASD. Two experimental protocols are designed to scan the spontaneous response data of the subjects in a non-intrusive manner.

The third goal addresses the challenge that the elicited facial responses of individuals with ASD can be too subtle for computing the differential traits. The subtle expressions in subjects with ASD may not be easily classified as one of the six prototypical expressions of healthy individuals. This requires sophisticated computational models to probe facial muscle-specific responses in a non-intrusive manner for the subjects with ASD. There appears to be no work in the literature that employs non-intrusive computational models to reliably investigate the facial muscle-specific responses of individuals with ASD. This dissertation performs two experiments to investigate the differential traits for ASD using non-intrusive and spontaneous facial imaging data. The first experiment uses the proposed Frenet frame-based computational model on 3D facial data to probe facial muscle-specific responses. The second experiment uses a sophisticated facial action coding system for the first time in the ASD literature to identify the differential traits from the prevalence of subtle facial actions. Complementary response data such as eye-gaze and percentage of correction recognition are considered to discover a behavioral marker for the group with ASD.

The fourth goal aims to investigate the time-sampled dynamic activations of different facial actions of individuals in response to dynamic social stimuli. An experiment is performed to capture and identify the differential traits using a non-intrusive facial motion capture system. This experiment offers a fine-grain analysis of facial responses to different contexts in audio-visual stimuli. This non-intrusive scanning of dynamic facial actions may be potential substitute for intrusive EMG-based scanning of faces.

In general, there is a significant gap in the literature in investigating computational models for faster, automatic, quantitative, and objective detection of differential traits for ASD. Although effective, the existing clinical procedures for screening ASD-related differential traits require significant time, cost, training, and expertise due to human interventions. The discovery of measurable traits via computational models may be used as behavioral biomarkers in the ASD screening procedures for the early detection of ASD-related deficits to enable an effective intervention strategy.

1.3 Contributions

The primary contributions for this dissertation are discussed below.

Following the overarching need for automatic, quantitative, and objective detection of differential traits in the facial expression of individuals with ASD, the first goal is to develop a novel 3D geometry-based computational model for generalized 3D face classification, recognition, and localization. For this purpose, a 3D Frenet frame-based computational model for analyzing 3D facial data is proposed. The proposed computational model with Frenet frame-based representation offers versatile benefits such as competitive performance in generalized face and facial expressions recognition, pose-invariant face recognition, data dimensionality reduction that results in faster processing of data, and localization of facial responses based on facial curves. The proposed model reduces the large volume of data in a 3D face to a small, pose-invariant, and effective feature space without compromising the performance in the recognition task. A mathematical proof as well as experimental demonstrations also show the pose-invariance properties of these features in pose-invariant recognition of face and facial expressions. The Frenet frame-based computational model also obtains location information about different facial regions through the curve-based representation of 3D face. This location information about facial curves contributes in two ways following their respective experimental results. First, following a feature selection and classification method, the curves corresponding to the selected features can be identified and located on the 3D facial data. Second, the location of a specific curve can be targeted over the sample of faces to probe actions related to a particular facial region. This localization attribute of the proposed computational model is later considered to probe the responses at different facial muscles using the 3D facial response data collected from the subjects with ASD. The proposed computational model is evaluated using three publicly available benchmarked datasets of 3D face and facial expressions.

The second goal addresses the limitations in the existing psychophysical studies to obtain a non-intrusive and ergonomic experimental setup for data collection for subjects with ASD. Institutional Review Boards (IRBs) from two different institutions (ODU and EVMS) approved two separate carefully designed psychophysical experiments. The study recruited a total of 36 human subjects, a half of whom have a diagnosis of ASD and the other half includes typically-developing (TD) subjects. The first experiment collects data on the spontaneous psychophysical responses using a video camera, 3D facial imaging sensors, and an eye-tracker. The second experiment employs a motion capture sensor to capture the dynamic facial actions in response to dynamic audio-visual stimuli. Both

experiments involve a hardware-software platform that integrates multiple non-intrusive imaging sensors for data collection, synchronizes the sensor activation with the onset of the visual stimuli presented to the human subjects, and automates the data collection and the sequential procedures for the experimental protocols. The remaining goals and contributions are proposed following the data collected from these two experiments.

The third goal is to investigate the proposed 3D curve-based computational model in the quantification of a behavioral marker for ASD namely the 'oddity' in the facial expressions of individuals with ASD. The differential trait in the facial responses of subjects with ASD is reported as oddity in the qualitative behavioral studies on ASD. Statistical analyses in this dissertation show significantly asymmetric responses in a pair of facial muscles for the group with ASD. This suggests a quantitative metric for computing a differential trait in a non-intrusive manner from the geometric features of the facial curves.

The fourth goal is to discover behavioral markers from the facial responses captured by video data, 3D facial images, and visual responses from the eye-tracker data. For the first time in the ASD literature, the facial action coding system (FACS) is introduced to classify spontaneous facial responses using video data and 3D facial images. The differential traits of the subjects with ASD are studied from the prevalence of facial action units (FAUs) defined by FACS. State-of-the-art computer vision frameworks are proposed to recognize the prevalence of different FAUs in the facial responses of subjects with ASD. FACS-based analysis contributes a better insight into the subtle facial responses, which may not be visible to human eyes. The analysis of video frames of facial responses and eye-tracker data reveal behavioral biomarkers related to the facial expression and social smile for the group with ASD, which may be measurable targets in differentiating the group with ASD from the TD group.

The final goal aims to quantify the differential traits using the dynamic facial response of the subjects in response to dynamic audio-visual stimulus. In the second experiment, three metrics for computing differential traits are proposed from the prevalence of different FAUs in response to dynamic audio-visual stimuli. The first metric determines the prevalence of each FAU for each group, the second metric computes the correlation between time-sampled activation data for any two FAUs, and the third metric computes the stimuli context-specific activation of all FAUs for each group. These metrics facilitate a fine-grain analysis of the spontaneous facial response data to objectively evaluate the differential traits underlying the psychophysical responses of the subjects with ASD.

In summary, this dissertation proposes quantitative, non-intrusive, and objective computational models to investigate the differential traits from the spontaneous facial responses of subjects with ASD. This alleviates the limitations with the existing subjective and intrusive procedures in the literature involving posed or imitated facial expressions. The dissertation, so far, has contributed to four journal manuscripts and two conference proceedings. The computational model for analyzing 3D face using Frenet frame and its effectiveness in pose-invariant 3D face and facial expression recognition tasks are presented in a journal manuscript published in the IEEE Transaction on Human-and-Machine Systems (THMS) [21]. The curve-based localization of facial regions based in facial curves is published in a conference proceeding [22]. The findings related to the asymmetry in the facial responses of subjects with ASD are published in the Journal of Optics and Laser Technology (JOLT) [23] and in the 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM-2015) [24]. The findings related to the behavioral biomarkers for ASD are presented from the prevalence of FAUs and an eye-tracker in a journal manuscript currently under review in the IEEE transaction on Cognitive and Developmental Systems [25]. The results from the second experiment are presented in another journal manuscript, which is in preparation for submission in the IEEE transaction on Affective Computing [26].

1.4 Organization of the dissertation

The rest of the dissertation is organized as follows. Chapter 2 provides a background review on the differentials traits in ASD and state-of-the-art methods for investigating the traits from facial responses. Chapter 3 proposes the design and protocol for the two IRB-approved psychophysical studies. This chapter covers the criteria for human subject enrollment, the setup for the integration of the non-intrusive multimodal sensors, tasks, and the procedures for data collection. Chapter 4 investigates the efficacy of the proposed computational model for the analysis and classification of 3D facial data. Chapter 5 studies the computational model proposed in Chapter 4 to quantify the differential traits in the activation of facial muscles for the group with ASD in the first experiment. Chapter 6 proposes state-of-the-art computer vision models to compute prevalence of different FAUs from the video data of face and 3D facial data. The prevalence of FAUs is studied along with the eye-gaze patterns of the subjects to obtain a measurable trait as the behavioral biomarker for ASD. Chapter 7 proposes three metrics for computing the differential trait from the dynamic facial action data in response to dynamic audio-visual stimuli. The dissertation concludes with a summary and future work in Chapter 8.

CHAPTER 2

BACKGROUND REVIEW

The symptoms of ASD can be categorized into three major types. First, one of the signs of ASD is obsession or lack of attention that restricts an individual's ability to learn and adapt to social and behavioral development. More than one-third of the children diagnosed with ASD have some form of obsession [3]. Second, intellectual deficit is one of the co-morbid features present among 60% of children diagnosed with ASD [27], which may cause lack of proper perception of their surroundings and context. The lack of perception of the nonverbal communication skills such as human gesture, verbal cues, and facial expressions may limit the ability to learn necessary cognitive and motor skills for social and psychological processes. Third, repetitive and restrictive behaviors in individuals with ASD are mainly characterized by repetitive motor actions such as continuous hand flapping, lining up toys or spinning wheels of a toy car [3]. Subjects with ASD also restrict their interests to a limited area and stick to an action-based routine. This limits their ability to learn, adapt, and accept new aspects and situations in real life. These symptoms may lead to the manifestation of differential traits that can contrast a group with ASD from a group with typically-developing (TD) individuals.

2.1 Differential Study for ASD

The traits of ASD are identified in differential studies to compare a group with ASD to those of a TD group. The differential traits can be measurable, and hence, can be considered biomarkers for ASD [28, 29]. The differential traits in ASD have been investigated by scanning physiological responses and phenotype patterns using a variety of sensor modalities. There are prior studies that aim to find neural correlates of ASD from the brain fMRI data [30], MRI data [31], and evoked potentials in brain using EEG [32] as shown in Table 1. Liu et al. use several electrical sensors to collect physiological data for affective modeling of subjects with ASD [33]. Aldridge et al. report significant differences in the morphology of face between the ASD group and their control peers [34]. Apart from the facial morphology, other phenotypes related to head, hair, forehead [35], and face-brain asymmetry [36] have been considered to differentiate a group with ASD from the TD group.

TABLE 1: Application of different sensing modalities for quantitative differential studies on ASD.

References	Data modality	Data source	Detection
Aldridge et al. [34]	3D Camera	Face	Facial phenotypes
Liu et al. [33]	EMG EMG EEG	Two facial muscles, Shoulder, Cardiac response	Facial affect, Stress Anxiety, liking
Wang et al. [31]	MRI	Brain	Neural correlates
Deshpande et al. [30]	fMRI	Brain	Neural correlates
Thabet et al. [39]	CT scan	Brain	Auditory hypersensitivity
Wagner et al. [32]	Eye-tracker	Eye	Visual Area of interest
Volker et al. [17]	Manual Rating	Perceptual Skill	Perception about Facial expression

To address the attention deficits in the ASD group, prior studies quantify visual processing and attention of subjects from their eye-fixation patterns using an eye-tracker [32, 37, 38].

However, the verbal and nonverbal communication skills are the most common targets for clinical screening. A measurable trait or behavioral marker associated with the nonverbal skills has additional implications beyond contributing to objective diagnosis or screening procedures. Since ASD impairs more than one area of human actions and response to different extents, a measurable target can quantify the severity of the disorder. The measurement of psychophysical responses may identify the area of deficits to facilitate the subject-specific intervention strategy. The efficacy of an intervention can be objectively evaluated if the deficits can be quantified with the aid of a sensor. Therefore, the differential traits of ASD are necessary not only to learn about the deficits associated with the disorder but also to obtain a target metric for assessing the efficacy of intervention system [40]. The following sections describe the nonverbal behavior of subjects with ASD, which may offer potential targets for identifying differential traits of the disorder.

2.2 Nonverbal Behavior in ASD

Nonverbal response and communication provide important cues about behavioral engagement in a social setting. It has been suggested that nonverbal responses convey about 80% of the meaning in social communications [41]. Two of the most important channels of nonverbal communication are facial expressions and eye contact. Children with an effective use of facial expressions and eye contact are often adept at social interactions [42]. However, individuals with ASD have core deficits in processing mental states such as beliefs, intentions, desires, and emotional expressions during social communication [43]. Individuals with ASD often struggle to perceive human emotional cues such as facial expressions [17, 44, 45]. Volker et al. report that children with ASD have greater difficulty recognizing anger and fear expressions than their TD peers [17]. The failure to recognize social communications may, in turn, manifest atypical responses in nonverbal behavior.

Therefore, the study of facial expressions and eye movement data may identify useful behavioral markers for the group with ASD. The facial response in humans can happen following one of two mechanisms [17, 46]. First, it can be motor mimicry where an individual matches the expression as he/she visualizes the expression. This is a non-emotional mechanism to create facial responses. Second, a facial response can be spontaneous as a result of emotional arousal due to a stimulus. The spontaneous facial response provides an indication of the subject's perception and psychological development. For example, when an individual views an angry expression, the motor mimicry response will show anger. However, the response due to emotional arousal may yield a fear expression. There are some emotional expressions (e.g. laughing), which are typically contagious as they induce similar expressions [47]. There may be a presence of recurrent or atypical facial expressions such as in palsy, which may be subtle, uncontrolled, and not associated with a particular emotional state [48].

In the ASD literature, facial expressions are elicited using visual stimuli that are known to create unusual activations of certain brain regions in subjects with ASD [49]. This unusual activation may lead to atypical facial responses by altering the natural mimicry or empathy toward the displayed emotions. The facial expressions in response to visual stimuli may be complemented by eye-tracker data from an individual's visual scan. Eye-gaze is a pivotal element in determining an individual's engagement and focus of interest. While there has been some progress in understanding the visual scanning mechanism of individuals with ASD [50, 51, 32, 37, 38], there remain challenges in computing and encoding elicited and subtle facial responses in ASD.

TABLE 2: The state-of-the-art studies on the analysis of facial expressions of individuals with ASD.

Study	Modality	Non-intrusive	Spontaneous Expressions	Outcomes
[53]	EMG	No	Yes	Higher muscle activity in ASD
[16]	Visual Observation	Yes	Yes	Intense, but odd facial expressions in ASD
[17]	Visual Observation	Yes	No	Odd facial expressions in ASD
[54]	EMG	No	Yes	Undifferentiated
[19]	Facial markers	No	No	Less complex facial patterns in ASD

2.3 Facial Expression Analysis in ASD

The face is a sophisticated target that embodies motor attributes associated with emotions and gestures for nonverbal communication skills. A variety of facial muscles contribute to a diverse set of facial expressions that may reflect mental states, cognitive processes, mimicry of expressions, and other voluntary or involuntary motor actions triggered by the brain. Therefore, neurodevelopmental disorders like ASD may manifest the functional anomaly of the brain through the physiology of facial muscles. The Diagnostic and Statistical Manual (DSM) of Mental Disorders outlines the differential traits in eye-gaze and facial expressions during social interaction as two among other diagnostic criteria for ASD [52]. A controlled psychophysical study with proper visual stimuli can elicit responses in facial muscles, which may, in turn, unravel differential traits in the facial physiology of the individuals with ASD. The state-of-the-art psychophysical studies investigating the facial expressions of individuals are highlighted in Table 2, and can be categorized based on the following two categories.

2.3.1 Qualitative Study

In one of the earliest studies on facial expressions of ASD, Adrien et al. conduct a retrospective analysis on home videos of children with ASD acquired from 11 families [55]. Their study, based on visual investigation, reveals differences in the facial expression and eye-contact between children with and without ASD. While physicians use visual observation and expertise during the diagnostic process, most of the research studies employ naïve human coders who qualitatively score the facial expressions from video data of the face [17, 16, 56]. Subjective rating scales are used to study the ability of individuals with Parkinson's disease in posing emotional facial expressions [57]. Stel et al. use an audiovisual stimulus to elicit facial expressions in subjects with ASD and employ naïve human coders to evaluate the facial expressions from the video data [56]. Unfortunately, human coder-based studies often suffer from poor inter-rater reliability. Volker et al. employ naïve human coders to evaluate the performance of participants with ASD in imitating prototypical facial expressions.The imitated facial expressions are compared with the target prototypical expressions to investigate anomalies in facial responses. Unlike spontaneous mimicry of emotion or empathy toward the viewed emotion, imitation is an effortful or volitional task [18, 58]. The task of imitation does not reflect the psychophysical response which is due to spontaneous emotional or social trigger.

In a similar retrospective analysis of facial images of individuals with ASD, Faso et al. report a more intense presence of facial expressions when compared with the TD group. However, such qualitative measurement of intense facial expressions reports less natural traits than those perceived in the TD participants. The lack of natural traits in facial expression is known as an odd facial expression, and such 'oddity' in the facial expression may be one of the differential traits of individuals with ASD [16, 17]. Although human eyes can perceive an emotion underlying a typical facial expression, neurodevelopmental disorders (e.g. ASD, Alzheimer's disease, schizophrenia etc.) may impair the facial expressions resulting in facial actions that are too subtle and ambiguous to trace and score visually. This necessitates computational modeling of the subtle facial expressions to assist the clinicians with objective differential information.

2.3.2 Quantitative Study

The existing quantitative techniques have mostly utilized electrophysiological sensors in contact with several body parts including the face. Park et al. have performed a

quantitative study on the facial expressions of adult subjects with schizophrenia [59]. They employ multiple intrusive sensors to capture the psychophysical response of the subjects in response to emotion inducing video stimuli. The subtle facial responses in individuals with ASD have been computed with the aid of Electromyography (EMG) by placing electrodes at different facial muscle regions [46, 33, 53, 54]. Studies involving EMG measurements report either similar facial response with a time delay [60] or enhanced mimicry actions [61] in the group with ASD than those in the TD group. Unfortunately, this method limits the probes to only two facial muscles (*Zygomaticus Major* and *Corrugator Supercilii*) without considering the entire face [33, 46]. It may be infeasible to probe the entire face using such electrodes to yield a comprehensive profile of facial responses. Furthermore, the intrusive application of electrodes on facial skins may irritate or distract an individual with ASD in different ways, since half of the population diagnosed with ASD is known to have anxiety and fear of novel experiences as well as tactile defensiveness [3]. The application of EMG electrodes involves intrusive steps such as washing the face with soap, preparing facial skin using alcohol swabs, and finally placing Ag/AgCl electrodes on the facial skin. Electrophysiological sensors have been also placed on different body parts such as the face, fingers, shoulder, and chest of individuals with ASD [33]. In such a constrained environment, the subjects are often restricted in their body, hand, and head movements. This may bias natural facial and visual responses of the subject to the visual stimuli. In order to yield unbiased and natural psychophysical responses, recent quantitative studies on ASD avoid wearable or intrusive sensors on any body parts [62, 17] and employ computer vision-based methods for the analysis of behavioral traits.

2.4 Computer Vision-based Study

The availability of sophisticated optical sensors has enabled non-intrusive scanning and analysis of the entire face for facial features in details. A recent computer vision-based study proposes a quantitative means to evaluate the severity of schizophrenia from the subject's ability to imitate different facial expressions [20]. As mentioned earlier, the imitation of facial expressions takes effort and hence, can be intense and deliberately exaggerated by an actor in achieving the target facial expression. In contrary, spontaneous facial responses in individuals with ASD due to expression-mimicry or emotional arousal can be subtle, ambiguous, and even attenuated due to the neurodevelopmental disorder. This poses an additional challenge for the computer vision-based methods in encoding and detecting the differential traits from subtle facial responses. Optical sensors, like 3D facial

cameras, have been employed to capture high-resolution 3D facial data of individuals with ASD in order to investigate atypical morphological features in the face [34]. Hashemi et al. have recently proposed computer vision-based methods to detect differential markers from the video data of body motion and gestures for the children at risk for ASD [62]. Unfortunately, there appears to be no work that applies computer vision methods to compute and compare the spontaneous facial responses in subjects with ASD with those in a TD group. The following section provides an overview of the state-of-the-art trends in the computer vision-based analysis of facial imaging data.

2.5 Analysis of facial imaging data

Facial imaging data from video data and from 3D imaging sensors as 3D facial point clouds are commonly utilized for the computer vision-based analysis and classification of facial attributes. The 2D intensity-based facial image frames from video data are the most common type of imaging modality. There have been an extensive studies that involve processing pipelines of 2D facial images for the retrieval of facial identity [63], facial expressions [64], and other face-related key-points and features [65]. However, 2D facial images suffer from a variety of non-trivial challenges such as facial pose [66], occlusion [67], lighting conditions [68], scale or size of the face [69], and so on. Therefore, a vast majority of the current literature on 2D facial images aims to propose robust algorithm pipelines to address one or more of these non-trivial challenges. A common method is to reconstruct 3D facial data from 2D images to mitigate the effects of these non-trivial challenges on the classification of 2D facial images [70].

On the other hand, a 3D representation of the face facilitates the analysis of the human face by providing more detailed geometric information than 2D intensity-based images. Unlike 2D intensity-based images, the classification of 3D facial data is less affected by distortions such as facial pose, occlusions, and illumination. These advantages have led to advances in 3D face recognition (3D-FR) [71, 72] and 3D facial expression recognition (3D-FER) [73, 74, 75]. Although 3D facial data are rich in geometric information, their classification entails increased computation due to high volumes of data. 3D facial data are most commonly reduced through manual selection of landmarks [76, 77, 73, 78] or 3D to 2D projection techniques [74, 76, 79]. These data reduction techniques may inherently limit the advantage of 3D over 2D data in pose invariant recognition of the face and facial expression. The following sections provide an overview of the related works on 3D-FR and 3D-FER. While 3D-FR is primarily a recognition task comparing two sets of 3D facial

data, 3D-FER is mostly a classification task among faces bearing different expressions.

2.5.1 3D Face Recognition

In a face recognition task, a probe face is compared with each facial image sample in a gallery of faces to determine the similarity in their facial features. The facial images in the gallery are ranked based on similarity score and the probe face is recognized as the gallery individual with the highest similarity score.

The goal of 3D-FR is to retrieve facial identity by finding global features that are invariant to local variations such as facial expressions [80, 81, 82], pose [80, 70], and occlusions [83] in facial images. Different global features preserving data reduction techniques such as average face, Eigen face [71], and isomap of the face [81] are less sensitive to facial expressions. The rigid region of the face is often selected to yield features insensitive to local variations [72, 84]. Sangineto proposed a method to localize facial features that are invariant to pose and expressions for 3D-FR [85]. Such approaches selectively exclude a part of the 3D facial data which may be useful for 3D-FER. In addition, such facial expression invariant 3D-FR approaches are effective under the assumption that deformation of the face due to expression is isometric [80, 81, 86], which may not be always the case [81]. Therefore, it is important to search for pose and expression invariant features from the entire face rather than arbitrarily excluding a facial region.

2.5.2 3D Facial Expression Recognition

The local variations in the face are the most important features for 3D-FER. The recognition of facial expressions is a multiclass classification task, where the number of classes is determined by the number of target facial expressions. When the expression is mild or occurs at the rigid part of the face, it contributes less to the local variation of the face [72, 84]. In addition, the facial shape of different individuals vary. Therefore, it is essential to extract effective local features that can represent different expressions independent of the global information of a face. To obtain representative features from 3D facial data, a common method is to project the 3D facial data onto 2D and use the 2D projected intensity-based features for classification. Previous studies used 2D projected geometric features such as curvature [74, 77], depth [74, 76, 79], curvature-based shape index [74, 77], and spherical harmonic features [87] for either 3D-FR or 3D-FER. However, such projected features lose the depth information of 3D data [88], and hence suffer from the limitations that are common in 2D face classification. Wang et al. report that the

recognition performance of facial expressions from such 2D projected features of 3D facial data rapidly degrades with increasing angle of the facial pose [76]. Therefore, the facial features should be extracted directly from the 3D facial data without entailing loss of data as in 3D to 2D projection.

2.5.3 Recognition of Facial Actions from Images

Facial action coding system (FACS) provides a dictionary of all possible facial muscle actions encoded in Facial Action Units (FAUs) [89, 90]. A prototypical facial expression (e.g. happy, angry) can be synthesized from multiple FAUs. Conversely, this enables a fine-grain analysis of facial expressions by decomposing a facial expression into multiple of the constituent FAUs [20, 91]. Therefore, subtle and ambiguous facial expressions can be decomposed and categorized into units of facial actions known as FAUs. The analysis or synthesis of FACS-based FAUs has been contributing to the areas of human-computer interfaces, multimedia and animation applications, security and surveillance, and behavioral research. Behavioral research studies aim to learn differential traits in psychophysical responses of a target group under a certain task or stimulus. The psychophysical responses of subjects with neurodevelopmental disorder such as ASD can be subtle and ambiguous and may not be classified as prototypical facial expressions as seen within the TD subjects [20]. Therefore, the analysis of facial expressions based on FACS may provide a better insight into the subtle psychophysical responses of subjects with ASD, which may not be evident to human eyes. The FAUs defined by FACS can also serve as markers for different emotional states. For example, upper lip raising (FAU 10) is found in disgusted facial expressions [92]. Lip corner puller (FAU 12) represents a smile [47]. The facial dimple, contributed by the buccinator muscle, yields FAU 14, which is reported as a marker of depression [93]. Lip pressing is annotated by FAU 24, which represents resentment. 'Lips apart' is defined by FAU 25, which appears in the surprised expression. Figure 1 [1] shows the spatial location of different facial muscles contributing to the activation of different FAUs.

Within recent literature, Wu et al. [94] and Savran et al. [74] provide two of the most comprehensive studies on the recognition of FAUs from 2D facial image frames and both 2D and 3D facial images, respectively. Wu et al. investigate an optimal design of a multi-layer feature extraction architecture for the best possible recognition of FAUs. This work suggests that a dual layer feature extraction architecture where Gabor features are

---

[1]Collected from: http://www.kidport.com/reflib/science/HumanBody/MuscularSystem/
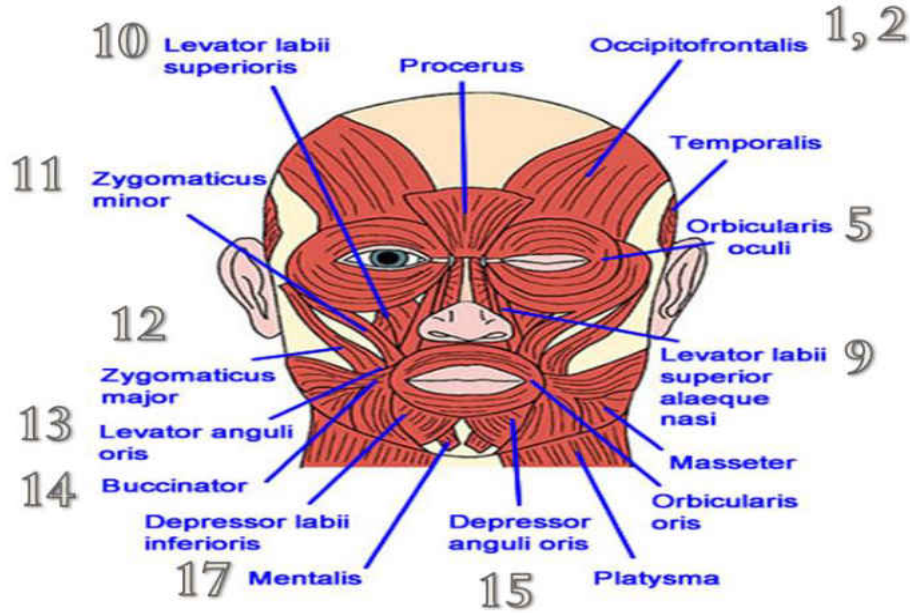
FIG. 1: Spatial locations of different facial muscles contributing to the activation of different facial action units using FACS[1].

extracted in the first layer and Local Binary Patterns (LBP) in the following layer yields better recognition accuracy than using any one of the layers alone. Savran et al. perform an extensive research on the performance of recognizing different FAUs from 2D and 3D facial images. They develop the Bosphorous dataset that provides both 2D and 3D facial images labeled with different FAUs. This work concludes that 3D facial images provide better accuracy than 2D images in recognizing a number of facial action units. However, 2D facial images do outperform 3D images in recognizing a few other FAUs. The computer vision-based detection of FAUs recently shows a fine-grain analysis of only imitated facial expressions in subjects with schizophrenia [20]. FACS-based computational models may play significant role in the analysis of spontaneous facial expressions of subjects with ASD. This will require a computational model capable of recognizing the prevalence of multiple FAUs from the images of spontaneous facial expressions.

2.6 Challenges in Analyzing 3D Facial data

Although 3D facial data offer more detailed information about the face than 2D intensity-based images, there remain several challenges in analyzing 3D facial data. For

example, the pose invariant classification of 2D facial images has been an intense area of study [70, 78]. However, the effect of pose distortions in 3D face classification has not received much attention until recently. Similar to 2D facial image, the capture of 3D facial data also suffers from the loss of information due to an out-of-plane rotation of the head. This loss of information can deteriorate the performance of 3D-FR and 3D-FER. The degree of pose angle variations in 3D facial data poses an additional challenge in effective feature extraction for classification. Such pose related distortions may be mitigated by extracting facial features directly from 3D facial data. For 3D-FR, Lei et al. use distance and angle features obtained from 3D facial vertices [72], as well as the depth information at different point locations of 3D facial curves [95]. Ocegueda et al. compute the probability of each vertex of 3D facial data to classify faces for 3D-FR and 3D-FER [96]. Such methods mostly involve multilayer networks and heavy computations for extraction and classification of features from the large volume of 3D facial points. These 3D point-based methods often involve a computationally expensive point-to-point registration step to align two 3D facial data samples. As an alternative, representative landmark points are often selected manually from the 3D facial data [77, 73, 78]. Wang et al. decompose the 3D face into three 2D image maps each representing one of three components of 3D normal vectors computed at the vertices of 3D facial surface. They manually segment the 2D projected 3D facial features into 21 regions for 3D-FER [97]. Such manual segmentation or selection of landmarks is often cumbersome and limits the automated FR [80]. In addition, these methods may discard a significant region of the face which may be useful for a recognition task. Youssef et al. report that the scanning of the entire face provides better 3D-FER performance than selecting only the upper or lower half of the face [98]. Therefore, it is necessary to take the entire face into account to ensure robust 3D-FR and 3D-FER. These necessitate an effective computational model that can offer automatic extraction of pose-invariant features from the intrinsic geometry of the 3D face.

2.6.1 Curve-based Analysis of 3D Face

The geometric features are extracted from point [71, 96], or surface [74], or curve-based [88, 80] representation of 3D facial data. A curve-based representation can be a good trade-off between surface and point-based representations for computational efficiency. A surface-based representation of a 3D face provides useful global features for 3D-FR at the expense of high computation cost [88]. A point-based representation, on the other hand, provides local information which may be more effective for 3D-FER than a surface-based

representation [72]. However, the registration or comparison of 3D facial data in large datasets using 3D points can be computationally more expensive when compared to that of surface-based methods [96, 99, 71]. Unlike surface or point-based representation, radial curves can be easily localized, parameterized, and compared.

There are studies that involve the curve-based representation of 3D face for 3D-FR [83, 88, 99, 80]. Li et al. compare curves from six facial regions to perform 3D-FR [83]. Jahanbin et al. use facial curves extracted from different facial contour paths to compare 3D facial data in 3D-FR [88]. Samir et al. represent 3D facial data into a collection of level curves to compare 3D facial data from the differential geometry of the curves [99]. However, these studies involve an additional 3D pose correction step for 3D-FR. Drira et al. [80] report that radial curves offer benefits over other curve-based representations for 3D-FR. Radial curves, originating from the nose tip, are easy to acquire and useful to probe different facial regions. This dissertation proposes radial curve-based representation of 3D face to facilitate an efficient classification of face and facial expressions and to enable analysis of subtle facial responses from the geometry of the curve.

In summary, the detailed geometric information of 3D facial data can be useful in the analysis of subtle and spontaneous facial expressions when the challenges with 3D facial data are addressed. This requires sophisticated computational modeling to extract useful features efficiently, which can reveal differential traits for the group with ASD. On the other hand, an efficient psychophysical study protocol and non-intrusive methods for facial data acquisition are necessary to elicit the facial responses of subjects with ASD. This will facilitate computer vision-based computational models in discovering the measurable differential traits from the facial expressions of individuals with ASD.

CHAPTER 3

EXPERIMENTAL DESIGNS FOR ASD STUDY

3.1 Chapter Overview

This chapter discusses the design and protocol for the two experiments conducted in this dissertation examining the facial responses of individuals with ASD. As discussed in Chapter 2, the experimental protocol and setup need to carefully consider the behavioral and physical manifestations of individuals with ASD. An ergonomic experimental setup is required that offers minimally or non-intrusive access to body parts, as well as facilitates maximum engagement without any distraction. The stimuli need to be appropriately chosen to elicit responses without causing any extreme experience such as shock or surprise. The response data from the participant should be collected over a fixed time frame synchronous with the visual stimulus in order to make the data comparable between subjects and trials of the study. All these factors should be taken into account while designing a controlled experimental setup. The study protocol has been designed following multiple studies [46, 100, 101, 56] and in consultation and collaboration with experienced physicians at Eastern Virginia Medical School (EVMS) and an expert in communication disorders at Old Dominion University (ODU), who are actively engaged in providing health services to children with ASD.

3.2 Subjects

The study protocols for the two experiments, involving human subjects, are approved by the Institutional Review Boards (IRB) at ODU (See Appendix A) and EVMS, respectively. In the first experiment, a total of sixteen human participants have been recruited as follows: (i) a group exclusively diagnosed with ASD with IQ score >70 and (ii) a healthy or typically developing group (control group) without a diagnosis of ASD. The age range of the participants is between 7 to 20 years which is suitable to undertake the tasks for the study. Among sixteen enrolled subjects, eight are diagnosed with ASD by a psychiatry practice at EVMS with an average age of 13±4.4 years. The remaining eight subjects are typically-developing (TD) individuals with an average age of 16±4.1 years. The ages of the
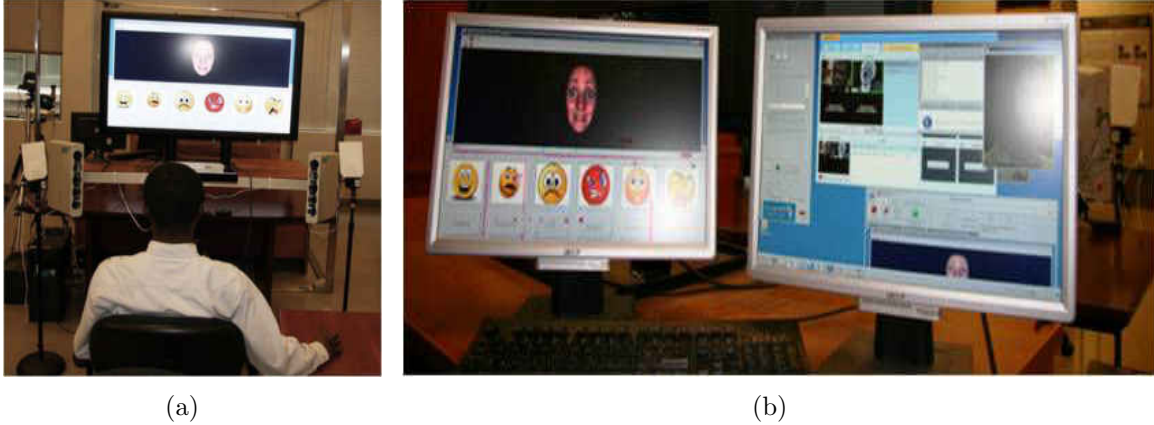
FIG. 2: Hardware-software setup for the experiments. (a) Subject sitting before the 68"
TV display and the positioning of the camera sensors. (b) Two monitors behind the scene
to launch and coordinate the experiments.

two groups are compared under an ANOVA test and the results show an F-value of the test
of $F(1, 14)=2.66$, and a large p-value ($p>0.05$). Such results are indicative of no significant
difference in the ages between the two groups. All the subjects without ASD are considered
within the TD group. For the TD group, the same inclusionary and exclusionary criteria
(See Appendix B) are followed except the diagnosis of ASD or any disorder is considered
as an exclusion criterion. In the second experiment, the same protocol is followed to enroll
a total of twenty age-matched subjects. The ASD group consists of ten subjects with
average an age of 13.5 ± 2.37 years. The TD group includes ten subjects with an average
age of 13.1 ± 3.31 years.

Prior to the enrollment, a potential subject is screened for inclusion and exclusion
criteria via telephone. An eligible subject and parent undergo the informed consent and
assent forms, which briefly outline the goals and tasks in the study.

3.3 Hardware-Software Platform

A hardware-software platform is developed to facilitate an ergonomic accommodation
of the subjects, as well as an automatic operation of the sensors and stimuli. The platform
includes a 68" multimedia TV to display the visual stimulus as shown in Fig. 2(a). Lerner
et al. place their subjects 2.5 feet before a 24" monitor [100], which is transformed to
seven feet for a 68" TV display. There are also two monitors behind the scene as shown
in Fig. 2(b). The experimenter uses these two monitors to launch and coordinate each

(a)

(b)

(c)
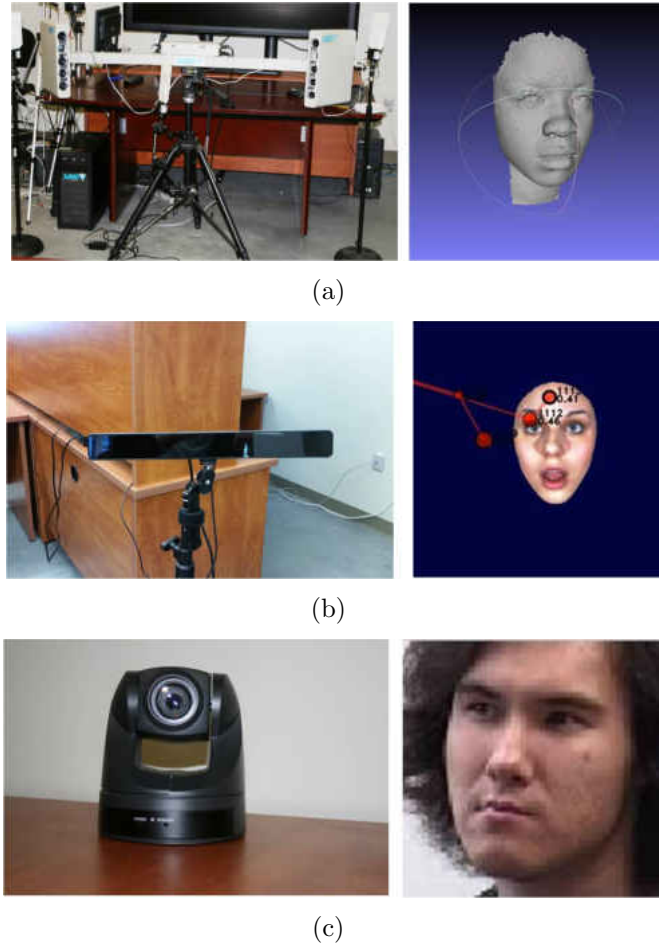
FIG. 3: Three non-intrusive sensors and their outputs for visualization. (a) 3dMD camera for 3D facial imaging, (b) Mirametrix eye-tracker showing the eye-gaze map, (c) Sony EVI-D70 video camera for videotaping of the face.

trial during the study. One of the monitors clones the 68" TV display and the other displays the command and control applications for the experiment. The non-intrusive sensors used in this study are as follows. A commercially available state-of-the-art 3D optical camera *3dMD* (www.3dmd.com). The 3dMD system is reported to reconstruct 3D point cloud geometry from the captured stereo images with an average error of 0.27 mm [102]. This eliminates chances for any significant error in the 3D data analysis due to the 3D reconstruction process. The video data of the face is captured using a Sony EVI-D70 color video camera. An eye-tracker camera from *Mirametrix* (www.mirametrix.com) is also placed in the setup to capture the time-sampled data about the gaze location and duration of the participants. Fig. 3 shows the sensors and corresponding generated data
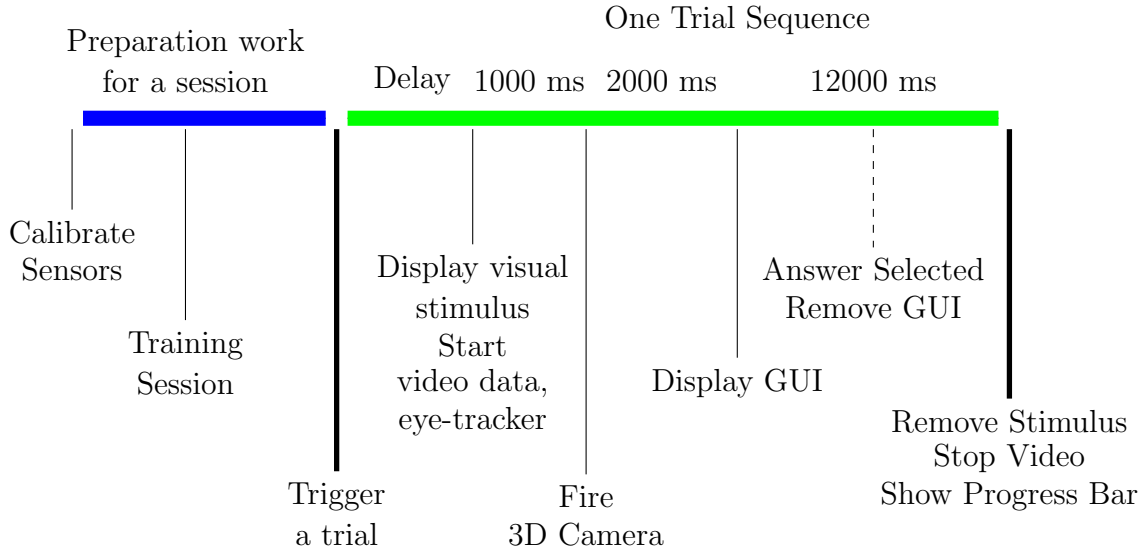
FIG. 4: Preparation work for a session and timing diagram for automated sequence of actions in a trial of the study for the first experiment.

used in the first experiment.

All the sensors are positioned away from the line of sight of the participants as they view the TV screen. For each subject, the eye-tracker and the cameras are calibrated to have the subject's face within the field of view. A graphical user interface (GUI) application is developed using Qt/C++, which displays options for selecting an emotion from six buttons corresponding to six facial expressions. The GUI collects the subject's perception of the emotion displayed as a visual stimulus in each trial. The onset and end of a trial are programmed to happen at a definite time sequence as shown in Fig. 4. The mouse cursor is programmed to automatically and sequentially initiate each step in a trial, which involves automatic activation of the imaging sensors, display of the visual stimulus and the GUI at definite time instances. The timing diagram shows that once the GUI is displayed, the subject provides the answer by mouse cursor click. At the end of the trial, the cursor prompts the sensors to stop and remove the visual stimulus on time.

In the second experiment, a commercially available real-time facial motion capture (MoCap) system, known as *faceshift* (www.faceshift.com) is used to capture the dynamic facial motion of a subject in response to a dynamic audio-visual stimulus. The *faceshift* system uses a 3D range sensor named *PrimeSense* to capture the 3D depth image of the face. At first, the deformable 3D facial model with the *faceshift* system is calibrated by the

(a)   (b)
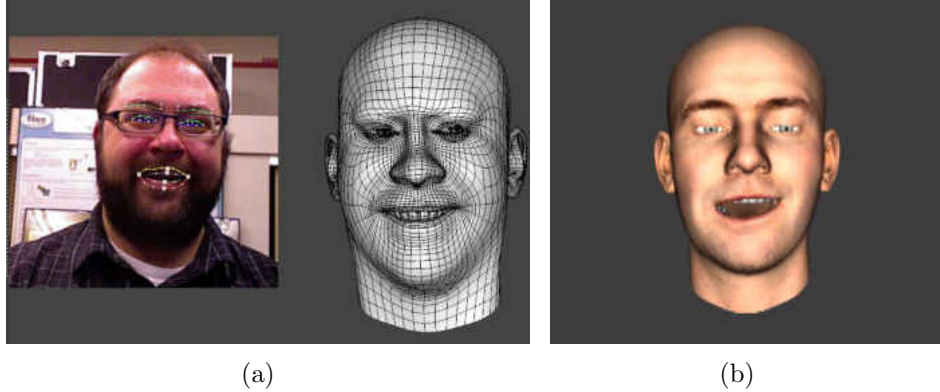
FIG. 5: Data collection in the second experiment. (a) Tracking of facial key points of the actor to map on the 3D deformable model for rendering the stimuli and collecting subject data. (b) The animated Avatar's face used for the audio-visual stimulus.

neutral face of the subjects. The *faceshift* application tracks the facial landmark points (See 5(a)) that are computed in real-time from the 3D facial range image captured by the *PrimeSense* sensor. The motions of the facial landmarks are mapped onto the calibrated deformable 3D facial model to compute the magnitude of activations for different facial actions. The magnitude of facial actions is normalized between the values of 0 and 1. In this experiment, the *faceshift* facial MoCap system is used to develop the Avatar-based audio-visual stimuli, as well as to collect the spontaneous time-sampled magnitudes of the facial actions of the subjects in response to the audio-visual stimuli. Table 3 summarizes the subject and design information related to the two experiments. Fig. 7 shows the application of *faceshift* MoCap system in rendering the audio-visual stimuli and in collecting facial action data from the subjects for the second experiment.

3.3.1 Stimuli

The first experiment uses high-resolution static 3D facial images with expressions (See Fig. 6) as visual stimuli, which are selected from the Binghamton 3D facial expression dataset [103]. It is hypothesized that visualization of such facial images portraying emotional expressions will either create mimicry or arousal of facial expressions [46]. In this study, the visual stimuli are collected from the Binghamton 3D facial expression dataset that consists of high-resolution 3D faces benchmarked with six prototypical expressions (happy, anger, fear, sad, surprise, and disgust) [103]. The use of the 3D facial images has introduced more realism in the visual stimuli than the traditional 2D facial images with

TABLE 3: Subjects and designs for the two experimental studies.

	ASD Num., (Age)	TD Num., (Age)	Dura-tion	Stimuli	Tasks	Sensor Data
Expt. 1	8, (13±4.4) years	8, (16±4.1) years	40-min	Static 3D faces with Expression	Visualize, recognize, manipulate	3D face, video, Eye-tracker
Expt. 2	10, (13.5±2.3) years	10, (13.1±3.3) years	15-min	Audio-visual animation	Visualize, engage	Facial motion capture

(a)   (b)   (c)   (d)   (e)   (f)
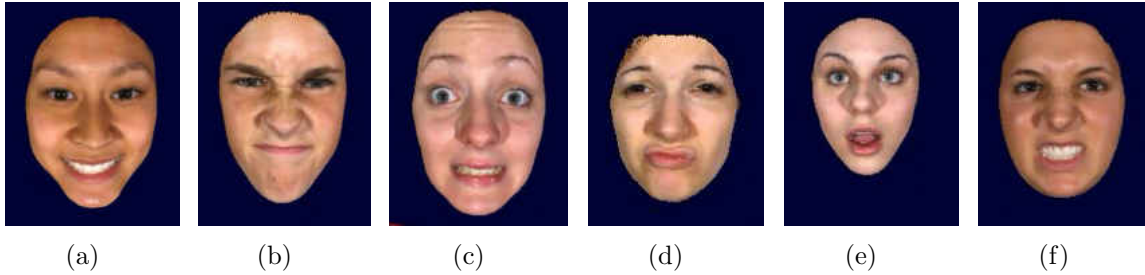
FIG. 6: Sample 3D facial images with emotional expressions (from the Binghamton 3D facial dataset) randomly displayed in different trials of the first experiment. 3D faces portraying (a) happiness, (b) anger, (c) fear, (d) sadness, (e) surprise, and (f) disgust expressions.

expressions used in [46, 100].

The second experiment uses an audio-visual stimuli [54, 56]. Individuals with ASD have been found to be more engaged during computer simulated environments and with Avatar characters than during real-person interactions [104]. The application of Avatar-based intervention in the ASD literature has been promising [105, 106], which motivates the use of Avatar-based stimuli to elicit spontaneous facial responses in this study. The persona of a real human actor is used to create an Avatar character via the *faceshift* application as shown in Fig. 7(Top). The Avatar character imitates the actor's speech and facial actions in a story-telling scenario, which talks about experiences related to his
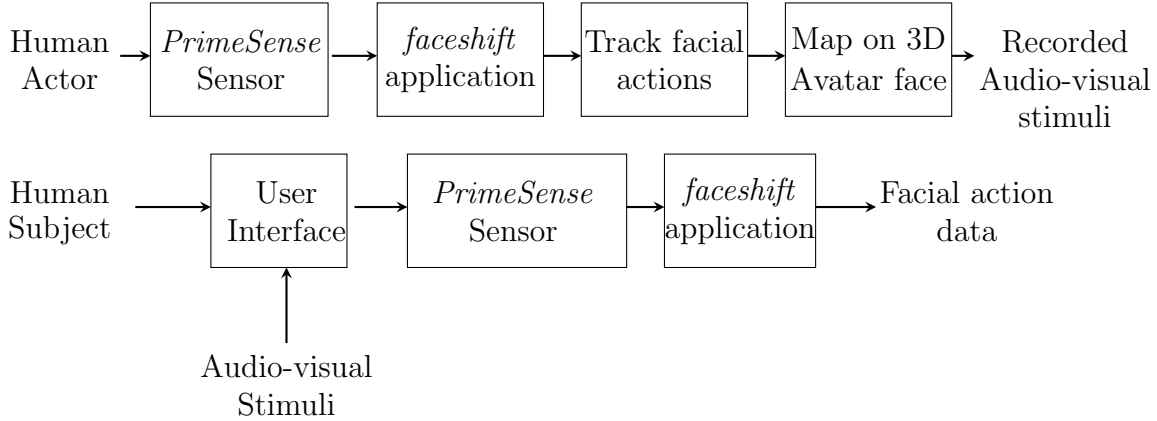
FIG. 7: Data acquisition steps for the second experiment. Top: Rendering of Avatar-based audio-visual stimuli mapping the persona of an actor. Bottom: Collecting facial activation data of subjects in response to audio-visual stimuli.

school-life story and engagement as narrated in Appendix C. The content and context of the story are selected at the appropriate interest and intelligence level of the target age-group. The speech and facial animation of the Avatar character (See Fig. 5(b)) are recorded as a four minute video data to be used as the audio-visual stimuli for the second experiment.

3.4 Tasks and Procedures

The subject is seated in an adjustable chair in front of a 68" multimedia TV with the computer mouse on a side table. In the first experiment, there are two sessions of tasks as shown in Fig. 8. The first session requires the subject to recognize the emotion from a visual stimulus of 3D face with expression in 12 random trials. A similar approach to elicit facial expressions in subjects with ASD is found in [46, 100]. Six facial expression samples with the highest expression intensity are considered from the dataset to create 12 random trials with 12 different 3D faces. Each trial displays a 3D face corresponding to one of six basic facial expressions in a random order. The experimenter triggers each trial independently to launch a sequence of automated actions. Following Lerner et al. [100], the subject is allowed to visualize the stimulus for 3000 ms before displaying the GUI to respond (Fig. 8 (a)). Furthermore, a time constraint of 15s is imposed for each trial [101] to allow the subject select from the GUI about the displayed 3D facial expression as shown in the timing diagram (See Fig. 4). The GUI disappears as soon as the answer is noted or

(a)	(b)
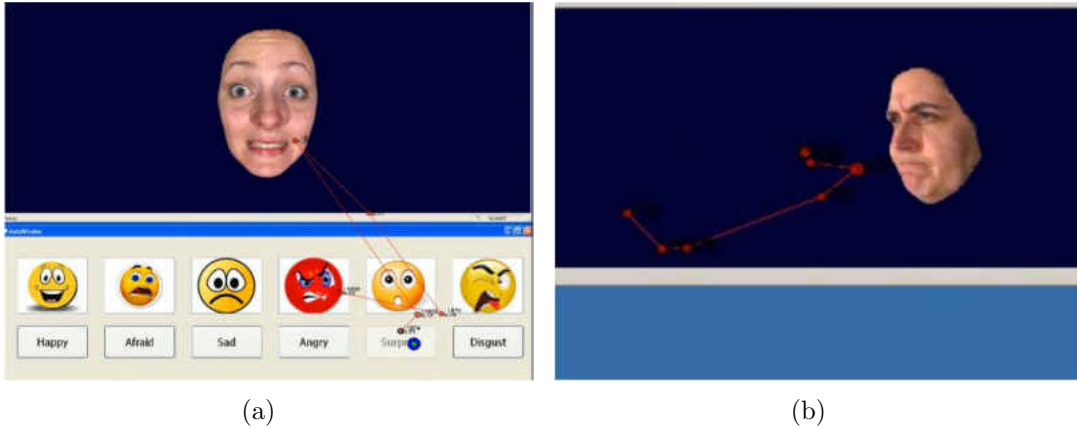
FIG. 8: Stimuli and tasks for the first experimental study. (a) First session: Facial expression recognition task from 3D facial expression stimuli, the colorful GUI at the bottom has buttons to click and choose about the displayed expression. (b) Second session: Manipulation of 3D face with expression using the mouse cursor.

TABLE 4: A summary of the two sessions for the first experimental study.

Session	Duration per trial	Num. of Trials	Display	Tasks
1	15 s	6 Expressions X 2 =12 3D-faces	GUI followed by 3D face	Visualization, expression recognition
2	25 s	6 Expressions X 2 =12 3D-faces	3D faces with expressions	Visualization, manipulation

at the end of 15,000 ms time constraint.

In the second session, the subject performs a manipulation task using a mouse in each trial as shown in Fig. 8(b). A total of 12 random trials are launched with 12 random 3D faces with expressions. The time required for the manipulation task may vary depending on the difficulty of the task. Based on observation with some test subjects, the subject in the study is given 25,000 ms in each trial for the manipulation task. At the end of a trial in either session, the mouse cursor automatically replaces the visual stimulus with a progress bar and a note of encouragement. The subject performs several practice trials prior to each session. A summary of the two sessions for the first experiment is presented

in Table 4. The subjects are given break periods in between two sessions or at any time they prefer. It takes about 40 minutes to complete the study including two sessions, calibration, positioning, and break periods. In the second experiment, the *PrimeSense* sensor is adjusted on a tripod to capture the facial image without interfering with the line of sight of the subject. The task of the subjects is to visualize, listen, and recognize the story content narrated by an animated Avatar.

3.5 Data collection

Several studies report that the facial affect appears 500-1000 ms after the onset of the visual stimulus [46, 54]. Therefore, 3D optical and the video cameras are activated 1000 ms after displaying the stimulus as shown in Fig. 4. The 3D camera takes about 14 s to charge its flashes before it can snap the next shot. Therefore, this camera yields only one 3D facial image sample per trial. The 3D camera yields dense 3D point clouds of the subjects face in response to the visual stimuli. On the other hand, the video camera continuously captures facial image frames over the entire period of the trial until it is terminated by the automated mouse cursor. Unlike static 3D facial imaging data, the video data offer a dynamic response of the face over the tasks of the trial that include visualization, recognition, and manipulation with the mouse cursor. The time-sampled 2D facial images from the video data are considered for analysis. Similar to the video camera, the eye-tracker captures the dynamic visual response data in the form of gaze fixation time and fixation location. The time-sampled data about the gaze fixation provide quantitative information about the visual interest and engagement of the subject. A note of encouragement with a progress bar is displayed at the end of each trial to lessen the anxiety and keep the subjects informed of their progress. In the second experiment, the facial actions of the subjects are collected via the *faceshift* application in response to the Avatar's story-telling and animation as shown in Fig. 7 (bottom). The faceshift application provides the time-sampled magnitude of a number of facial actions normalized between the values of 0 and 1. The onset of the facial data collection, the initiation of the audio-visual stimuli, and stopping of the data collection are synchronized and performed automatically. This allows a time synchronization between the presentation of the visual stimuli and the facial response data collected during the study. Upon completion of the study, each subject is offered a $15 gift card as an incentive for the participation as approved by the IRBs.

In the following chapters, novel computational methods are proposed to extract differential traits for the group with ASD from the multimodal facial images and eye-tracker

data collected in the first experiment. Statistical analyses are also performed on the dynamic facial action data collected in response to dynamic audio-visual stimuli in the second experiment.

CHAPTER 4

COMPUTATIONAL MODELING FOR THE CLASSIFICATION OF 3D FACIAL DATA

4.1 Chapter Overview

This chapter proposes a novel mathematical framework and its computational implementation for extracting curve-based geometric features from the 3D facial point cloud data. The aim is to exploit the inherent advantage of processing 3D data for pose invariant 3D-FR and 3D-FER, respectively. Without resorting to 2D projection, novel geometric features are extracted from the 3D space curve representation of 3D facial data. The Frenet frame-based geometric features are inherently fast to obtain when compared to the existing 3D point and surface-based representations. The proposed framework effectively reduces a large volume of 3D facial data into a finite number of radial curves and extracts a generalized set of novel geometric features from radial curves for a recognition task. The Frenet frame-based mathematical framework is used to represent the large volume of the 3D facial data in a pose-invariant feature space of very small dimension. The proposed framework and features alleviate the need for an expensive registration step that is usually needed in the 3D-FR pipeline.

4.2 Proposed Method for 3D Face Analysis

Following the discussions in Sections 2.6 and 2.6.1, this dissertation proposes a curve-based feature extraction method from the entire 3D face to facilitate a framework for 3D-FR and 3D-FER. This section illustrates the proposed framework to obtain pose invariant geometric features and their evaluation methods in 3D-FR and 3D-FER with and without pose distortions. The proposed framework, from pre-processing to experimental design, is discussed below.

4.2.1 Preprocessing of 3D facial data

A 3D face may contain non-facial regions, outliers, holes, and uneven surface. A surface smoothing technique, which applies averaging over the 3D vertices, is used to smooth the coarse facial surface and remove the surrounding outliers. An automatic preprocessing step is applied to clean the undesired components to yield a useful region of the 3D face. The nose tip is simply detected as the highest z-value of a 3D facial point cloud. The nose tip coordinate information is used to create a 3D mask equivalent to an average 3D facial region. The 3D mask crops the non-facial regions to obtain a valid 3D face. Finally, an interpolation step fills out any holes on the facial surface.

4.2.2 Normalization of 3D facial data

The 3D facial point cloud is nonuniform in 3D space. This point cloud can be represented as a surface, z=f(x,y) where z represents depth values as a function of nonuniform sample points (x, y). A normalization step interpolates all 3D facial data on a square grid with uniform sample points (x', y'), which is defined within the following boundary values,

$$[x_{min}, x_{max}] \quad = \quad \{x' \in R : x_{min} \leq x' \leq x_{max}\}, \tag{1}$$

$$and [y_{min}, y_{max}] \quad = \quad \{y' \in R : y_{min} \leq y' \leq y_{max}\}. \tag{2}$$

Here, x_{min}, x_{max}, y_{min}, and y_{max} are the boundary values of the X-Y coordinate for a given 3D face. The non-uniform points z=f(x, y) of a given 3D face are used to compute the interpolated values z' for the uniform sample points on (x', y') grid such as $z' = f(x', y')$. The interpolated values on the uniform grid provide a common topology for curve and feature extraction.

4.2.3 Extraction of radial curves

Once the 3D face is normalized following the interpolation step, the radial curves are extracted. Since frontal 3D faces are posed along the Z-axis, the nose tip can be automatically detected as the largest Z-value of the face as in [81]. The nose tip is projected as the origin (x_n, y_n) of the uniform grid. With the nose tip at the origin, the grid is divided into four quadrants. A set of linear paths (Fig. 9 (a)), defined by slope ratios $\frac{p}{q} = \frac{y_n - y_u}{x_n - x_u}$ in different quadrants, are originated from (x_n, y_n) and can be represented as,

$$y - y_n \quad = \quad \frac{y_n - y_u}{x_n - x_u}(x - x_u). \tag{3}$$
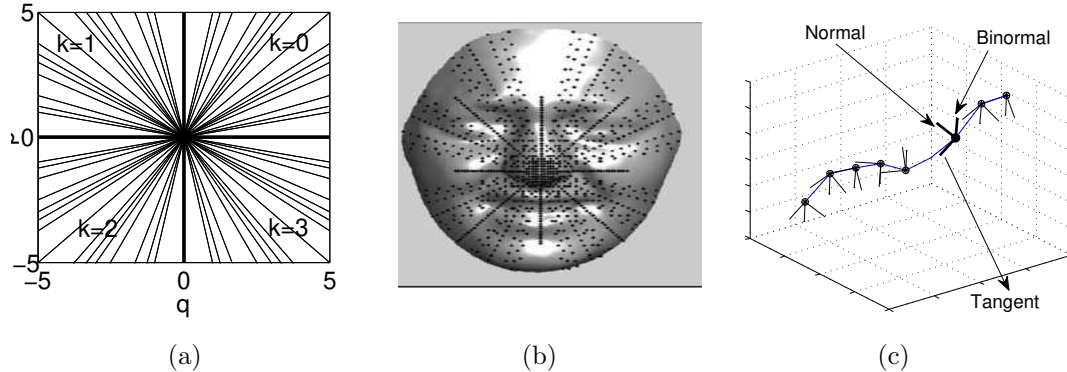
FIG. 9: Radial curve and Frenet frame representation. (a) Illustration of 64 linear paths with the nose tip at the origin. p and q determine the slope and the value of k selects the quadrant; (b) Extraction of 64 radial curves from 64 linear paths; (c) Frenet frame with 3 orthonormal basis vectors: Tangent $\hat{T}$, (Principal) Normal $\hat{N}$, Binormal $\hat{B}$ at each point of a space curve.

Here, (x_u, y_u) are the point coordinates on the uniform grid intersecting the linear paths. The interpolated depth values z_u, corresponding to these (x_u, y_u) points, constitute a 3D curve on the facial surface. Thus, the linear paths in Eq. 3 can obtain facial curves originating from the nose tip as shown in Fig. 9(b). These curves are known as radial curves and can be defined using an angular representation, θ_{ij}, given as,

$$\theta_{ij} = \frac{k_i\pi}{2} + \arctan\left(\frac{p}{q}\right)_j, k = 0, 1, 2, \text{and } 3, \tag{4}$$

where four values of k denote four quadrants and $\frac{p}{q}$ is the slope ratio for an arbitrary linear path j in a quadrant.

4.2.4 Generalization of radial curve using Frenet frame

The proposed mathematical framework for the curve-based geometry is adapted from [107]. Each of the extracted radial curves is considered as a space curve. At each point of the curve, a local coordinate system consisting of three orthogonal basis vectors can be constructed as shown in Fig. 9(c). This coordinate system is known as a Frenet frame [108]. A space curve is defined as $\beta : (a, b) \rightarrow \Re^3$ within an interval (a,b). In the Cartesian coordinate of $\Re^3$, $\beta(s)$ is defined as, $\beta(s) = (x(s), y(s), z(s))$, where x(s), y(s),

z(s) are real valued functions of real variables and s is the arc length parameter. A closed interval [a,b] of the curve can be partitioned in m vertices as:

$$B = \{a = s_0 < s_1 < \ldots < s_m = b\}, \tag{5}$$

where the position vector of a point f on a space curve $\beta(s)$ is $\vec{r}(s)$=(x, y, z). A neighboring point g on the curve corresponds to $\vec{r}$(s+ds). The tangent vector to the curve $\beta(s)$ at point f, $\vec{T}(s)$ is given as $\frac{d\vec{r}(s)}{ds}$, where $d\vec{r}(s) = \vec{r}(s) - \vec{r}(s + ds)$. The tangent vector at point f to the curve $\beta(s)$ can be expressed by:

$$\beta'(s) = lim_{\Delta s \to 0} \frac{\beta_{s_i} - \beta_{s_{i-1}}}{\Delta s}. \tag{6}$$

Here, $\|\beta'(s)\|$ is called the derivative of arc length at a point f to the curve, where the arc length is defined by $\int \|\beta'(s)\| ds$. From the definition of tangent vector given above, allows for the computation of the unit tangent vector $\hat{T}$:

$$\hat{T} = \frac{\beta'(s)}{\|\beta'(s)\|}, \text{ where} \tag{7}$$

$$\beta'(s) = \|\beta'(s)\| \hat{T}. \tag{8}$$

The derivative of Eq. 8 with respect to the arc length s yields another vector of the curve, β'':

$$\beta'' = \|\beta'\|' \hat{T} + \hat{T}' \|\beta'\|. \tag{9}$$

By definition, a vector of a constant length and its derivative are orthogonal to each other. Since, $\left|\hat{T}\right| = 1$, $\hat{T}'$ is orthogonal to $\hat{T}$ and also orthogonal to its principal normal vector, $\hat{N}$, such that, $\hat{T}.\hat{N} = 0$. Since, $\hat{T}'$ is proportional to $\hat{N}$, Eq. 9 can be rewritten as follows:

$$\beta'' = \|\beta'\|' \hat{T} + \|\beta'\|^2 \kappa \hat{N}, \tag{10}$$

where, κ is the curvature at a point on the space curve. In $\Re^3$, the third normal vector, the binormal vector, $\hat{B}$, is such that, $\hat{B} = \hat{T} \times \hat{N}$. The triple $\{\hat{T}, \hat{N}, \hat{B}\}$, as orthonormal basis vectors, create the Frenet frame field of β as shown in Fig. 9(c). Both β' and β'' vectors of a curve β are represented in terms of the orthonormal basis vectors of Frenet frame. The cross product of these vectors can be expressed as follows:

$$\beta' \times \beta'' = \kappa \|\beta'\|^3 \hat{B} \tag{11}$$

Taking the magnitude on both sides in Eq. 11 yields $\|\beta' \times \beta''\| = \kappa \|\beta'\|^3$. Here, $\hat{B}$ is a unit normal vector which yields the expressions for κ:

$$\kappa = \frac{\|\beta' \times \beta''\|}{\|\beta'\|^3} \tag{12}$$

Replacing the κ in Eq.11 by Eq. 12, an expression for unit binormal vector ($\hat{B}$) can be obtained as follows:

$$\hat{B} = \frac{\beta' \times \beta''}{\|\beta' \times \beta''\|}. \tag{13}$$

Similarly, the derivative of β'' as β''' with respect to arc length s can be expressed in terms of Frenet frame basis vectors to yield the expression for torsion (τ) as below.

$$\tau = \frac{\beta' \times \beta''.\beta'''}{\|\beta' \times \beta''\|^2}. \tag{14}$$

4.2.5 Pose invariance of Frenet-based generalized space curve features

The space curve vectors (β', β'', and $\beta' \times \beta''$) related to basis vectors of a local coordinate frame are not invariant to motion in 3D space. The basis vector field $\{\hat{T}, \hat{N}, \hat{B}\}$ will transform with the facial pose which is a form of Euclidean motion in $\Re^3$. Here, the projections of β', β'', and $\beta' \times \beta''$ vectors are shown along these basis vectors of a Frenet frame are invariant under Euclidean motion or pose angle of the face. The projection of β' along an orthonormal basis $\hat{T}$ is $\|\beta'\|$ as shown in Eq. 8. Applying an affine transformation matrix A on β' yields γ', given as $\gamma' = A\beta'$. We take the magnitude of this expression, $\|\gamma'\| = \|A\beta'\|$. Since the determinant of an orthogonal transformation matrix is one, $|A| = 1$, we can write $\|\gamma'\| = \|\beta'\|$. Therefore, $\|\beta'\|$ is invariant to motion in Euclidean space in $\Re^3$. Similarly, from Eq. 11, the projection of $\beta' \times \beta''$ vector along orthonormal basis vector $\hat{B}$ is $\kappa \|\beta'\|^3$. $\kappa \|\beta'\|^3$ can be also invariant to Euclidean motion if κ is invariant. The transformation $\gamma' = A\beta'$ can be applied in Eq. 9 to obtain an expression for β'', which is, in turn, plugged in Eq. 12 to obtain $\kappa[\gamma] = \kappa[\beta]$. Hence, κ is invariant to transformation in 3D space. Similarly, from Eq. 14, it can be shown that $\tau[\gamma] = \pm\tau[\beta]$, and hence τ is also invariant.
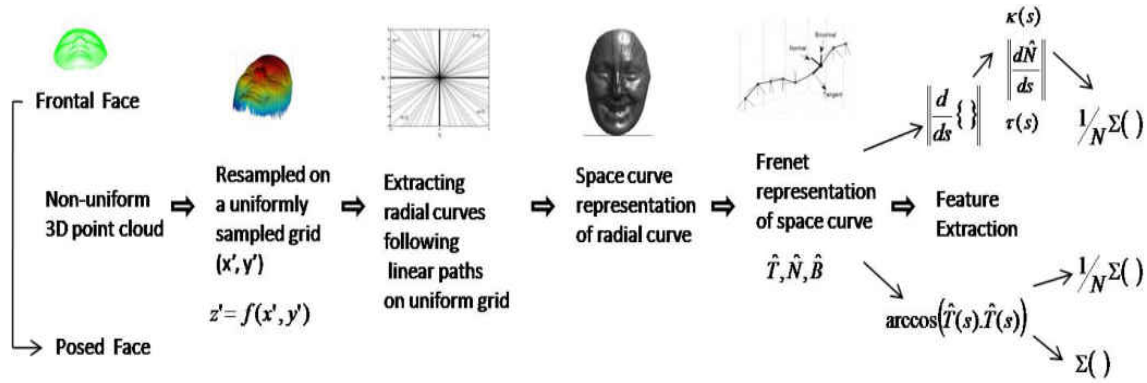
FIG. 10: Algorithm flow diagram for the mathematical framework of feature extraction.

4.2.6 Feature Extraction

Fig. 10 illustrates a summary of the proposed Frenet frame-based feature extraction framework from 3D facial data. We derive five types of features from the Frenet frame representation of space curves as follows. The derivatives of these basis vectors $\left\{\hat{T}, \hat{N}, \hat{B}\right\}$ with respect to arc length parameter s can also be represented as linear combinations of the basis vectors using Frenet-Serrete formulae [109] given as,

$$\hat{T}' = \kappa\hat{N}, \tag{15}$$

$$\hat{N}' = \tau\hat{B} - \kappa\hat{T}, \text{ and} \tag{16}$$

$$\hat{B}' = -\tau\hat{N}. \tag{17}$$

The magnitude of the derivative of basis vectors yields the desired new set of pose-invariant features in Euclidean space as follows,

$$\left\|\hat{T}'\right\| = \kappa, \tag{18}$$

$$\left\|\hat{N}'\right\| = \sqrt{\kappa^2 + \tau^2}, \text{ and} \tag{19}$$

$$\left\|\hat{B}'\right\| = \tau. \tag{20}$$

The magnitude of change of basis vectors, as shown in Eqs. 18-20, are computed at all points on each curve. For each curve, these values from all points are averaged to obtain the mean curvature, F_κ, mean torsion F_τ and the average change in principal normal vector,

F_n, respectively as shown in Eqs. 21-23.

$$F_\kappa = \frac{1}{N} \sum_{i=1}^{N} \kappa(s_i), \tag{21}$$

$$F_\tau = \frac{1}{N} \sum_{i=1}^{N} \tau(s_i), \text{ and} \tag{22}$$

$$F_n = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\kappa(s_i)^2 + \tau(s_i)^2}. \tag{23}$$

Mean curvature and mean torsion are represented by the average change of tangent and binormal vectors to characterize how the space curve turns and twists along its path, respectively.

The remaining two features are derived from the pose invariant feature, $\|\beta'\|$. For point-wise partition of a space curve as in Eq. 5, we compute the piece-wise arc lengths between each pair of consecutive points from their corresponding tangent vectors β'. Consider that each space curve $\beta(s)$ lies on the Riemannian manifold. The distance on the Riemannian manifold, between two consecutive points indexed by i and j on a curve $\beta : [i, j]$, can be measured by taking the dot product of their tangent vectors. From the definition of Reimann space, the arc length ds can be obtained using $ds^2 = \beta'_i.\beta'_j = \|\beta'\|^2$, which is invariant to motion in Euclidean space. The angle between two tangent vectors is α, where $cos\alpha = \frac{\beta'_i.\beta'_j}{|\beta'_i||\beta'_j|} = \beta'_i.\beta'_j$, since $|\beta'| = 1$. If α is the angle between a circular path with the radius as the unit tangent vector, the arc length of the path is equal to α. Therefore, we can write the piece-wise arc length (δ) of a space curve as follows,

$$\delta = \arccos\left(\beta'_i.\beta'_j\right). \tag{24}$$

Consider the sum of piece-wise arc lengths to yield the total length of the space curve is as:

$$F_\delta = \sum_{i=1}^{N} \arccos\left(\|\beta'(s_i)\|^2\right). \tag{25}$$

The mean arc length of a curve F_γ can also be measured by averaging the piece-wise arc lengths of the space curve as:

$$F_\gamma = \frac{1}{N} F_\delta. \tag{26}$$

Input: 3D Facial Data
Output: Frenet frame-based features
LUT: Lookup Table with slope ratios $\frac{p}{q}$ about the linear path
Nose tip: (x_n, y_n)
for i =1:num_of_curves **do**
    $\frac{p}{q} \leftarrow$ LUT (i)
    Linear path on grid $(x_u, y_u) \leftarrow$ {Nose tip, $\frac{p}{q}$}
    Radial curves $\beta_i(x_u, y_u, z_u) \leftarrow$ Calculate z_u for (x_u, y_u)
    $\{\hat{T}, \hat{N}, \hat{B}\} \leftarrow$ Compute Frenet frame for β_i
    X (i) $\leftarrow$ Compute features from $\{\hat{T}, \hat{N}, \hat{B}\}$; using Eqs. 21-23 and 25-26
end for

FIG. 11: Algorithm for the Frenet frame-based feature extraction.

Each curve is represented by one of five features as defined in Eqs. 21-23 and 25-26. The feature extraction algorithm is shown in Figure 11. Each feature value represents a curve which is obtained by either summing or averaging the features from all points on the curve. Therefore, a change in the local features will have less of an effect on the average feature values. This may facilitate expression invariant 3D-FR. On the other hand, the averaging operation eliminates the effect of the large number of points and corresponding feature values. If a larger facial shape yields a longer curve with more points per curve, the averaging may minimize the effect of an individual's facial shape to provide more local information required for 3D-FER.

4.3 Experimental protocol

To evaluate the generalizability and pose invariance of the proposed framework and features, experiments have been conducted in 3D-FR and 3D-FER for both frontal and varying pose conditions. For the evaluation in 3D-FR, two publicly available datasets (Face Recognition Grand Challenge (FRGC v2.0) [110] and Binghamton University 3D Facial Expression (BU-3DFE) [103]) are used. For 3D-FER evaluation, two 3D facial expression datasets (BU-3DFE and Bosphorous 3D facial expression [111]) are used. Figure 12 shows samples of 3D faces with expressions from these two datasets. Also, pose distortions are synthetically applied on 3D facial data to evaluate the performance for pose invariant 3D-FR and 3D-FER.
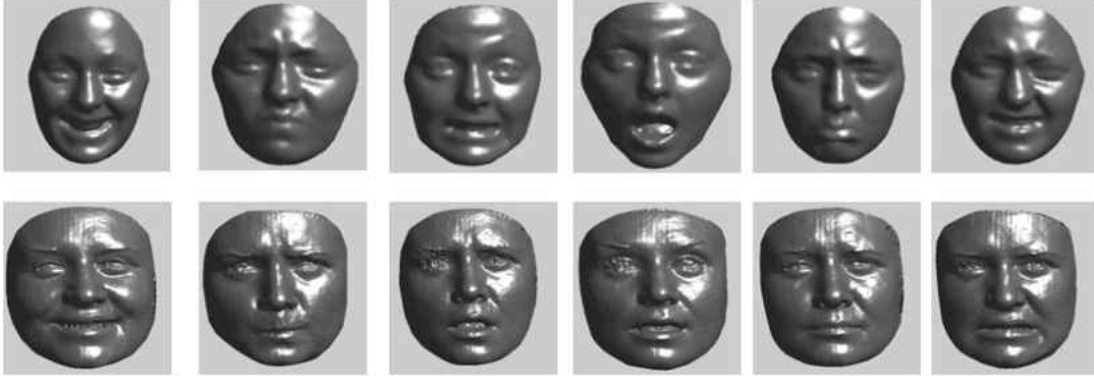
FIG. 12: Sample 3D facial expression data from the BU-3DFE dataset (top row) and the Bosphorous dataset (bottom row). Six expressions from left to right (happy, anger, fear, surprise, sad, and disgust) are selected for the experiment.

4.3.1 Experiment for 3D-FR

In the BU-3DFE dataset, the gallery consists of 3D facial data of different subjects with six basic facial expressions. The probe dataset consists of 3D facial data of the same subjects with neutral expressions. From the FRGC v2.0 dataset, the Fall 2003 dataset is used considering subjects that have 4 or more 3D face samples. For each subject, one sample of 3D face is reserved for the probe and the rest are considered in the gallery. The algorithm in Fig. 11 is used to extract Frenet-based features from each sample of 3D face.

To make the non-linear feature space separable for a classifier, three types of subspace projection techniques are used: Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Kernel Fisher Analysis (KFA). PCA projects the data onto the principal vectors which account for the highest variance in the data. LDA projects the data in such a way that maximizes the variance between-class data and minimizes the variance within class data. LDA is a generalized version of Fisher Discriminant analysis where the data distribution is considered normal. The kernel mapped feature applied to LDA is known as KLDA, which is interchangeably termed as Kernel Fisher Analysis (KFA). The features extracted from the gallery samples are used to train the corresponding model for subspace projection. The trained models are then used to project and select test features from the probe sample. The subspace projection facilitates a classifier such as nearest neighbor (NN) to effectively classify the projected probe features. The NN classifier is therefore used to train and test subspace projected features for 3D-FR.

The performance in 3D-FR is evaluated based on Cumulative Matching Curve (CMC)

and the Receiver Operating Characteristics (ROC) curves. For CMC, all samples in the gallery are ranked based on their similarity score with the probe sample. CMC shows the cumulative probability for the probe individual which can be correctly recognized within the top M ranked gallery individuals based on the similarity scores. Therefore, rank-1 recognition rate from the CMC is the probability for the probe individuals to be correctly recognized as the first or highest ranked gallery individual. For evaluation and comparison with ROC plots, the verification rate at 0.1% false acceptance rate (FAR) is considered.

When the test is conducted for pose distortion, the probe 3D face is rotated in an out-of-plane manner between 0^0 and 90^0 at 5^0 intervals prior to the feature extraction. With the increase of pose angle, the 3D facial point cloud gradually loses data in the direction of rotation. The effect of pose on 3D-FR performance is evaluated by the ROC curve and the area under the ROC (AuROC) curve corresponding to each pose angle.

4.3.2 Experiment for 3D-FER

3D facial point cloud samples with six basic facial expressions (happy, anger, fear, surprise, sad, and disgust) are considered from the BU-3DFE and Bosphorous datasets for classification. The algorithm in Fig. 11 is applied to extract Frenet frame-based features from each sample of 3D face. A multiclass feature selection technique, known as maximum relevancy and minimum redundancy (MRMR) [112], is employed to identify a set of representative features that best discriminate the 6-class facial expressions. The MRMR technique ranks the features under two combined criteria: One maximizes the mutual information between a feature space X and a target class Y and the other minimizes the redundant information considering the mutual information among the features. Both criteria are combined as follows,

$$\max_{m \epsilon X} \left[I(m, Y) - \frac{1}{|X|} \sum_{n \epsilon X} I(m, n)] \right], \tag{27}$$

where $I(m, Y)$ is the mutual information between feature m and target class Y and $I(m, n)$ is the mutual information between m and n features. A trial and error method is used to determine the dimension (N) of the features (the first N features from the rank) which yield the best accuracy in classifying the six facial expressions. The selected features are used to perform a 10-fold cross validation using three classifiers (Random Forest (RF), Naïve Bayes (NB), and SVM with radial basis function (SVM-RBF) kernel). Finally, the classification results are evaluated using ROC curves and AuROC. The approach, similar

<div align="center">(a)   (b)   (c)</div>
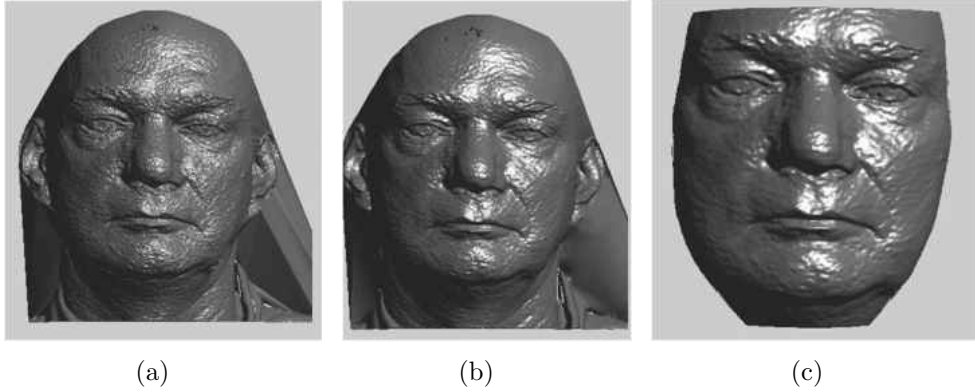
FIG. 13: Automatic preprocessing of a 3D face. (a) A raw 3D face with a coarse facial surface, (b) After smoothing the facial surface, and (c) After cropping the 3D face.

to 3D-FR, is applied to test the effects of pose distortion on 3D-FER. The 3D facial point clouds are rotated between 0^0 and 90^0 with an interval of 5^0 prior to the feature extraction. The extracted and selected features from the posed 3D facial data are tested for six facial expressions using a trained 6-class classifier model. The recognition rate versus the pose angle is plotted to show the effect of pose distortion on 3D-FER.

4.4 Results

This section illustrates the outcomes of the proposed framework and the results following the experimental protocols discussed in Sections 4.2 and 4.3.

4.4.1 Preprocessing

Following the preprocessing steps in Section 4.2.1, Fig. 13 illustrates an example of smoothing a 3D facial surface and removal of non-facial regions. The nose tip is automatically detected and subsequently used to define a 3D rectangular cuboid mask that retains the useful facial region after cropping the non-facial regions as shown in Fig. 13 (c). As discussed in Section 4.3.2, a 3D facial point cloud is normalized on a uniformly sampled square grid. However, the dimension of this uniform grid depends on the quality of the 3D point cloud rendering. For the BU-3DFE and Bosphorous datasets, a uniform grid of 100x100 is effective for a good recognition performance. For FRGC v2.0 dataset, the acceptable dimension of the uniform grid is found to be 200x200.
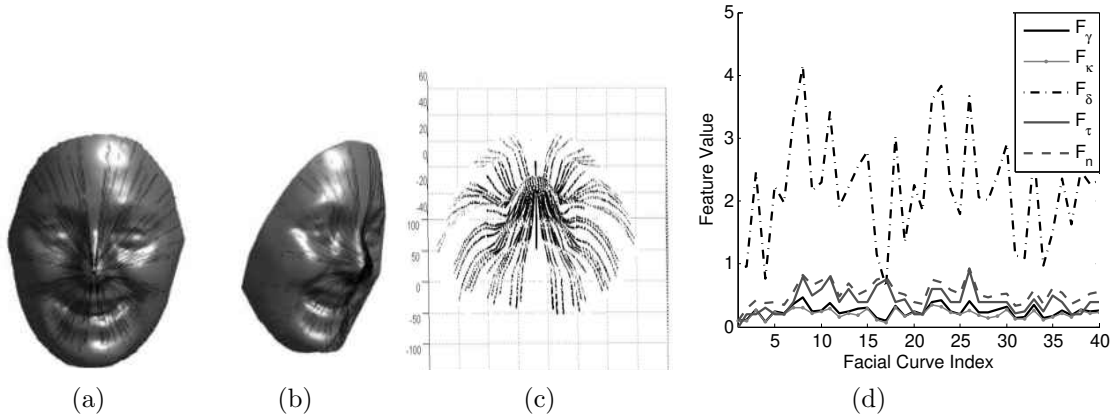
FIG. 14: Radial curves and features directly extracted from frontal and posed 3D faces. (a) 128 radial curves on a 3D frontal face; (b) Extraction of curves from a posed 3D face; (c) Topological view of extracted radial curves; and (d) Illustration of five types of features extracted from 40 radial curves.

4.4.2 Radial curve extraction

Radial curves, originating from the nose tip, are extracted following the procedures in Section 4.3.3. The same approach is applied for curve extraction from both frontal and posed faces without the need for a pose correction or a registration technique for the 3D facial data. Figure 14(a) and 14(b) show the radial curves on 3D frontal and posed faces, respectively. Figure 14(c) shows the extracted radial curves.

4.4.3 Feature extraction

The extraction of radial curves automatically decomposes a dense 3D facial point cloud into N number of 3D curves which are then directly transformed into an N-dimensional feature vector. Figure 14(d) offers a quantitative illustration of five types of features from 40 radial curves, where each radial curve is characterized by a feature value. This experiment suggests that a total of 192 radial curves per 3D face, with 48 curves per quadrant of the square grid, yields the best performance for both 3D-FR and 3D-FER. Unlike segmentation or selective landmark-based techniques, radial curves from the entire face are considered in search of important facial features which may discriminate facial identity and expressions.

TABLE 5: 3D-FR performance for different features and subspace projection techniques using the BU-3DFE dataset.

Frenet features	PCA		LDA		KFA	
	at 1% FAR	Rank-1	at 1% FAR	Rank-1	at 1% FAR	Rank-1
F_κ	72	74	82	86	90	92
F_γ	80	82	78	86	94	96
F_δ	72	82	72	80	84	94
F_τ	12	36	10	36	30	40
F_n	16	42	18	46	30	52

4.4.4 Results for 3D-FR

The BU-3DFE and FRGC v2.0 datasets were used to evaluate the performance of the proposed framework in 3D-FR as described in Section 4.3.1.

3D-FR using frontal faces

As mentioned earlier, for 3D-FR using the BU-3DFE dataset, the gallery contains 3D facial data with six facial expressions for each individual. The neutral probe faces are tested against the gallery of 3D faces with expressions. The features extracted from the gallery samples are used to create subspace projection models related to PCA, LDA, and KFA, respectively. These models are used to obtain a set of representative test features from the probe samples. Following the steps in Section 4.3.1., the NN based classification is applied on these test features to classify the probes for 3D-FR. For each type of feature and subspace projection technique, the ROC and CMC curves are obtained to provide the verification rate at 0.1% FAR and rank-1 recognition rate, respectively. Table 5 shows the summary performance for 3D-FR using different features and subspace projection techniques.

Table 5 shows that the F_γ feature outperforms the other four features and KFA provides the best 3D-FR performance. The first three features (F_κ, F_γ, and F_δ) show a comparable performance in 3D-FR with rank-1 recognition rates 92%, 96%, and 94%, respectively. In addition, Fig. 15 shows the CMC and ROC curves for 3D-FR with KFA-based subspace projection. The ROC curve in Fig. 15(a) shows the verification rates at 0.1% FAR and
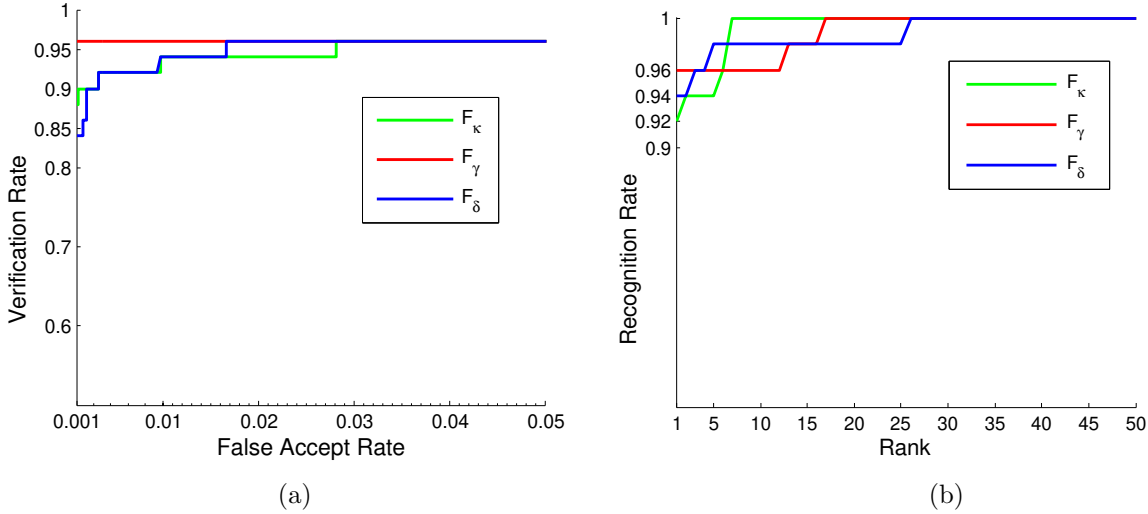
FIG. 15: (a) ROC curves and (b) CMC plots for three best performing features in 3D-FR using BU-3DFE dataset.

CMC plot in Fig. 15(b) reveals the rank-1 recognition rates for the three best types of features. Therefore, in the subsequent results for 3D-FR, the performance of these the best performing features (F_κ, F_γ, and F_δ) is demonstrated following the KFA-based subspace projection.

TABLE 6: Face recognition performance of different features and their fusions using FRGC v2.0 dataset.

Frenet features	LDA		KFA	
	at 1% FAR	Rank-1	at 1% FAR	Rank-1
F_κ	86	88	92	88
F_γ	88	92	94	96
F_δ	90	94	92	96
$F_\kappa + F_\gamma$	90	94	96	96
$F_\kappa + F_\delta$	88	94	92	96
$F_\gamma + F_\delta$	88	94	92	96
$F_\kappa + F_\delta + F_\gamma$	88	94	92	96

For the FRGC v2.0 dataset, the ROC curves and CMC plots in Figure 16 illustrate the 3D-FR performance of different features. Figure 16(a) shows verification rates at 0.1% FAR for three features (F_κ, F_γ, and F_δ) as 92%, 94%, and 92%, respectively. The rank-1 recognition rates from CMCs in Fig. 16(b) are 88%, 96%, and 96% for F_κ, F_γ, and F_δ,
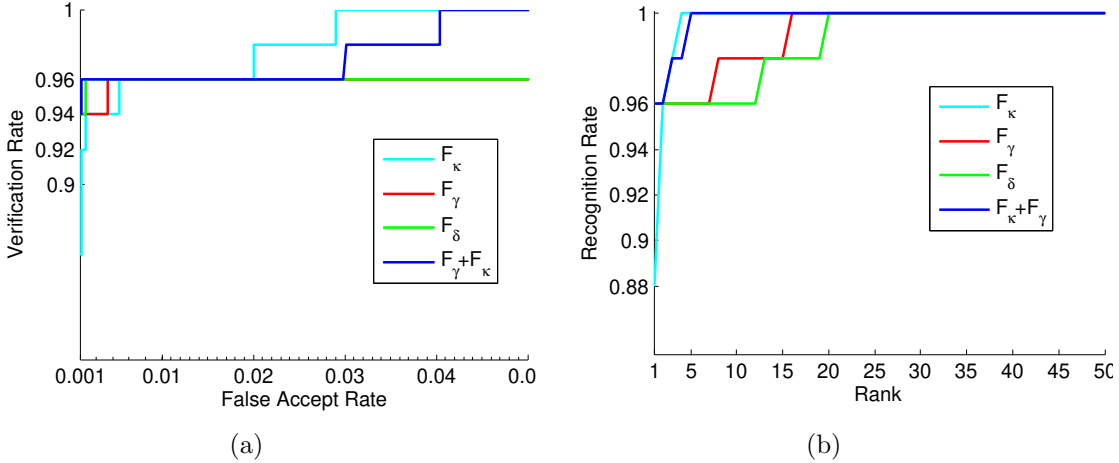
FIG. 16: (a) ROC curves and (b) CMC plots for the best performing features and a combination of features in 3D-FR using the FRGC v2.0 dataset.

respectively. Note that 3D-FR performance for the FRGCv2.0 is better than that for the BU-3DFE dataset. Since the BU-3DFE dataset contains a gallery of 3D facial data with different expressions, it may impose an additional challenge for expression invariant 3D-FR. On the other hand, the FRGC v2.0 dataset mostly contains 3D faces with neutral expressions. The absence of expression variations in the FRGC v2.0 dataset may have contributed to better performance than the BU-3DFE dataset. Further experiment was performed to see if any combination of these features provides any complementary information to improve the 3D-FR performance. A simple feature concatenation is performed using different combinations of the three best-performing features to probe their effectiveness for 3D-FR. Interestingly, the combination of F_γ and F_κ features yields the best performance in 3D-FR when compared to any of the individual features or their combinations. This combination yields 96% verification rate at 0.1% FAR and 96% rank-1 recognition rate, as shown in Table 6.

3D-FR using posed faces

In this experiment, the same gallery of 3D facial data, as in Section 4.3.1, is used to train the classifier. However, the probe 3D faces are synthetically rotated at different angles varying 5^0 between 0^0 to 90^0 prior to curve and feature extraction. The only assumption is that the coordinate location of the nose tip is known in the rotated 3D face to directly extract the radial curves and corresponding features. Figure 17 (a) shows
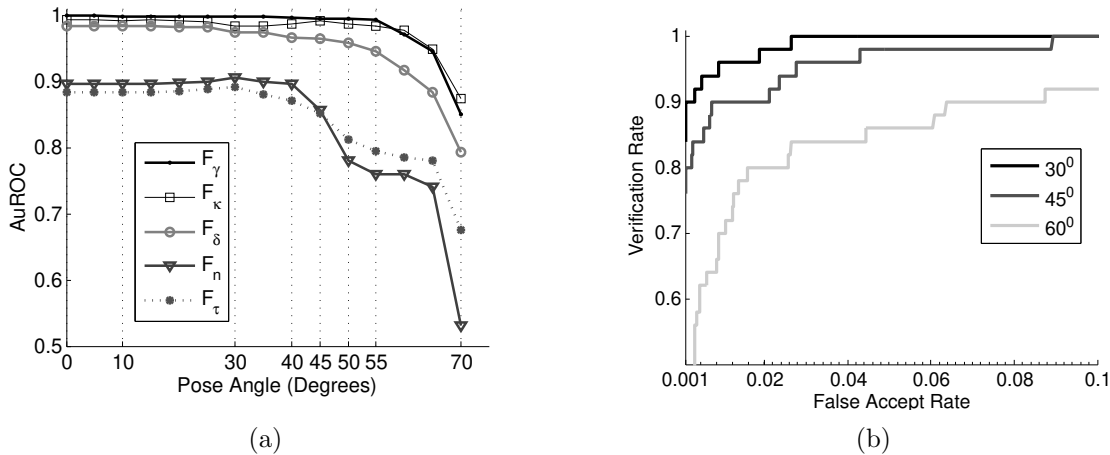
FIG. 17: Effect of pose angles on 3D-FR using the BU-3DFE dataset. (a) AuROC vs pose angles for five different features. (b) ROC curves for the best performing F_γ feature under different pose angles of the probe 3D face.

the 3D-FR performance for the five Frenet-based features extracted from 3D facial data at different pose angles while using the BU-3DFE dataset. Figure 17 (a) shows that the 3D-FR performance is consistent up to a 55^0 of pose angle for F_κ and F_γ, up to 40^0 for F_n and F_τ, and 35^0 for F_δ. F_κ and F_γ show the most robustness to pose distortion. Figure 17 (b) shows the effect of pose angle on the ROC curve for the best performing F_γ feature. The ROC curves demonstrate the decrease in 3D-FR performance with the increase in pose angle.

4.4.5 Results for 3D-FER

The BU-3DFE and Bosphorous facial expression datasets were used to evaluate the performance of the proposed framework in 3D-FER as described in Section 4.3.2.

3D-FER using frontal faces

The same curve and feature extraction framework, as shown in Fig. 11, is applied for 3D-FER. The three best performing features in 3D-FR (F_γ, F_κ, and F_δ) are also considered in this case for 3D-FER. The MRMR-based feature selection technique, as described in Section 4.3.2, is applied to obtain an optimal number of feature points for each type of feature, which yields the best classification of six facial expressions. Table 7 presents the AuROC for three features following a 10-fold cross-validation using each

TABLE 7: Area under the ROC (AuROC) curves after classification of the six basic facial expressions for three features using the BU-3DFE dataset. Three classifiers are RF = Random Forest, NB=Naïve Bayes, SVM = Support Vector Machine with radial basis function kernel).

Dataset	BU-3DFE								
Features	F_κ			F_γ			F_δ		
Classifier/ Expressions	RF	NB	SVM	RF	NB	SVM	RF	NB	SVM
Happy	95.6	93.9	97.0	96.6	94.2	96.7	95.9	93.1	97.1
Anger	85.4	80.2	83.9	88.6	82.5	85.5	87.9	83.2	84.3
Fear	79.8	81.8	79.8	83.2	83.2	81.2	83.4	82.5	81.6
Surprise	97.0	94.1	97.1	97.8	95.4	96.9	97.9	96.5	97.4
Sad	87.8	86.9	88.7	90.9	91.6	89.2	91.9	92	90.1
Disgust	91.6	90.6	91.5	91.1	91.0	90.7	89.6	88.9	90.0
Average	89.5	87.9	89.7	91.4	89.7	0.90	91.1	89.4	90.1

of the three proposed classifiers (Random forest, Naïve Bayes, Support Vector Machine) using the BU-3DFE dataset. The combination of feature F_γ with the multiclass random forest (RF) classifier offers the best performance in 3D-FER for both the BU-3DFE and Bosphorous datasets. Since the RF classifier is found to be the best performing classifier for the proposed feature type and dimension, this classifier is subsequently selected to evaluate similar features from the Bosphorous dataset as shown in Table 8. For both datasets, 'happy' and 'surprise' are found to be the best classified facial expressions. In comparison, the BU-3DFE dataset provides better 3D-FER performance than that of the Bosphorous dataset. This can be attributed to the fact that the BU-3DFE dataset offers better rendering of 3D facial data than the Bosphorous dataset [96]. In general, 'fear' and 'anger' expressions are poorly classified when compared to other expressions, which have affected the overall performance for 3D-FER. 'Fear' and 'anger' expressions are mostly characterized by eye and eye-brow regions at the rigid upper part of the face which do not contribute to the local variation similar to the mouth region [95, 84].

TABLE 8: Area under the ROC (AuROC) curves after classification of six basic facial expressions for 3 features using the Bosphorous dataset. Three classifiers are RF= Random Forest, NB=Naïve Bayes, SVM =Support Vector Machine with radial basis function kernel).

Dataset	Bosphorous		
Features	F_κ	F_γ	F_δ
Classifier/ Expressions	RF		
Happy	97.8	98	95.4
Anger	87.9	86	83.5
Fear	79.4	80.3	79.6
Surprise	88.4	87.9	85.7
Sad	77.5	77.5	75.1
Disgust	76.7	79.5	76.2
Average	84.6	84.9	82.6

3D-FER using posed faces

In this case, a multiclass SVM-RBF classifier model is trained using features from 3D facial data labeled with six facial expressions. The probe 3D faces are created by synthetically rotating the same 3D faces at different pose angles prior to curve and feature extraction. Features from these posed 3D faces are classified using the trained multiclass SVM-RBF classifier model. Figure 18 shows 3D-FER performance for three features under varying pose angles. For the BU-3DFE and Bosphorous datasets, F_γ yields the best performance as a pose robust feature in 3D-FER. For the BU-3DFE dataset, F_γ is pose-invariant up to a 55^0 pose angle as shown in Fig. 18 (a). F_κ and F_δ are pose-invariant up to 50^0 and 40^0 pose angles, respectively. For the Bosphorous dataset, F_γ is pose invariant up to a 55^0 pose angle while F_δ and F_κ are pose invariant to about 40^0. Figure 18 shows that F_δ is the most adversely affected feature due to pose distortion. This feature represents the sum of piece-wise arc length of the curves. With the increase of pose angle, the 3D face gradually loses data as well as the length of the curves. Hence, pose distortion affects the sum of arc lengths of the curve more than the mean of the arc length of the curve, F_γ.

4.5 Discussions

This work proposes a novel mathematical framework for Frenet frame-based generalized geometric features that can be effectively used in pose invariant classification of 3D faces.
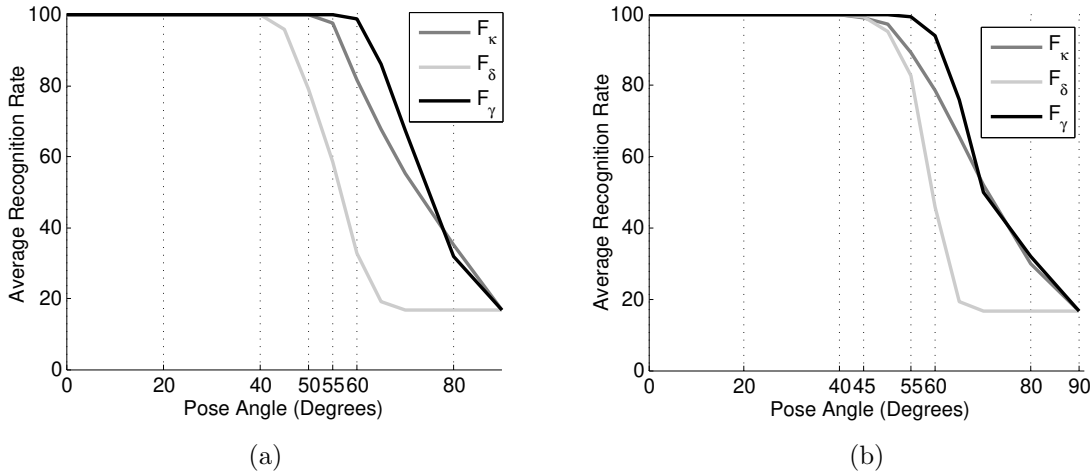
FIG. 18: Average recognition rate of six facial expressions for three different features under varying pose angles of test faces using the (a) BU-3DFE, (b) Bosphorous datasets.

The performance of the proposed features is evaluated in two recognition tasks: 3D-FR and 3D-FER with and without pose distortions.

4.5.1 Specific findings

The Frenet frame-based generalized space curve representation of 3D facial data can be effective for both frontal and pose invariant 3D-FR and 3D-FER. The mathematical and experimental results also verify the pose invariance property of these generalized features. The evaluation also suggests that the proposed features have varying performance for 3D-FR and 3D-FER. In general, the mean piece-wise arc length feature, F_γ performs the best for both pose-invariant 3D-FR and 3D-FER and can be considered as the most efficient generalized feature. However, the combination of F_γ and F_κ yields the best performance for 3D-FR with the FRGC v2.0 dataset. This suggests that these features can provide complementary information about the face. On the other hand, the sum of the piece-wise arc length feature F_δ may capture the local variations in the face image, which yield the facial expressions. Therefore, F_δ feature has a comparable performance with F_γ feature in 3D-FER as shown in Table 7. Overall, the F_γ, F_κ, and F_δ features outperform the other two features in both 3D-FR and 3D-FER. These three features are derived from the tangent vectors of the space curve. The contribution of the tangent vector suggests that the space curves from a 3D face bend like a circular path yielding a measurement of arc and curvature. The features derived from the other two vectors of Frenet frame, binomial and

TABLE 9: Comparison of 3D-FR verification rates at 0.1% FAR and Rank-1 recognition rates using the FRGC v2.0 dataset.

Method	Verification rate at 0.1% FAR	Rank-1 recognition rate
Alyuz et al. [113]	86.1%	not reported
Husken et al. [114]	86.9%	not reported
Al-Osaimi et al. [115]	94.1%	96.5%
Ocegueda et al. [96]	97.5%	96.6%
Faltemier et al. [116]	94.8%	97.2%
Drira et al. [80]	97.1%	97.0%
Dissertation Method	96.0%	96.0%

principal normal vectors, do not provide discriminative attributes for 3D-FR and 3D-FER. This finding suggests that the space curves from 3D facial data are not twisted along their paths in such a way that can be represented by these two vectors.

Under pose distortion, F_γ yields the best performance, which shows robustness to pose variance from 0^0 to about a 55^0 pose angle for both 3D-FR and 3D-FER. This indicates that for an out-of-plane pose distortion within 55^0 in one direction or 110^0 in both directions of the 3D face, the proposed feature may be able to provide pose invariant performance in a recognition task. Existing research employs a registration or similar step to correct the pose angle variation prior to 3D-FR [72, 80, 71], which can be computationally expensive for a large volume of 3D facial data. The proposed framework directly extracts the pose invariant features from posed 3D faces and eliminates the need for such a registration step.

4.5.2 Comparison with related work

Table 9 compares 3D-FR results using the FRGC v2.0 dataset. The results suggest either comparable or better 3D-FR performance for the proposed method when compared to the best performing methods. Similar 3D-FR performance with the BU-3DFE dataset in Table 5 suggests resilience of the proposed method to a large scale FR.

The proposed framework rapidly reduces about 30,000 3D points per 3D face down to an effective feature space with less than 200 feature points. The reduction in dimension retains the state-of-the-art performance in a synthetic biometric recognition task, as well as guarantees a faster processing time with 3D faces, which is crucial for real-time applications. Table 10 summarizes the superiority of the proposed method in terms of processing

TABLE 10: Cost of computation in registering and processing 3D facial data. ICP = Iterative Closest Point.

	Computation cost per 3D facial data in seconds			
	Registration	Processing	Total time	Processor
Li et al. [83]	1.95 (ICP)	7.05	9.00	Intel Core Duo 2.34 GHz
Fatlamer et al. [116]	2.38 (ICP)	7.50	9.88	Intel P4 2.4 GHz
Ocegadua et al. [96]	Pre-registered	2.88	>2.88	AMD Opteron 2.1 GHz
Dissertation Method	0 (Pose-Robust)	0.76	0.76	Intel Core Duo 2.66 GHz

TABLE 11: Comparison of 3D-FER % accuracy for six basic expressions using the BU-3DFE dataset.

	Vretos et al. [117]	Venkatesh et al. [118]	Gong et al. [79]	Dissertation Method
Happy	86.0	93.2	81.2	87.6
Anger	65.0	92.7	71.4	70.5
Fear	67.0	47.9	62.5	67.6
Surprise	89.0	89.9	88.1	87.6
Sad	62.0	74.0	77.5	76.2
Disgust	68.6	75.0	76.6	80.0
Average	72.9	78.8	76.2	78.2

time per 3D face. Although Faltermier et al. have a higher rank-1 recognition rate [116], their method takes 2.3 seconds to align a probe-gallery pair as shown in Table 10. This will take about 17 minutes to align a probe with 500 faces in a gallery, which is a limitation for real time 3D-FR. MatLab 2010b was used to run the algorithm in a 32-bit single thread process, which has the similar configuration for other studies. However, the magnitude of improvement in this work is considerably higher than all others.

Table 11 provides a comparison of the 3D-FER performance in this dissertation with those in the literature. The recognition of fear appears to be the most challenging. Tables 9 and 11 show that unlike existing studies, the same set of Frenet frame-based features

TABLE 12: Advantages of the proposed framework compared to existing methods using the BU-3DFE dataset.

Steps	Existing Approaches	Limitations of the Existing Approaches	Dissertation Approach
Feature Extraction	Manual [76, 77, 73, 78] Landmarks or Segments	Tedious, not automatic [80]	Automatic, Generalized features, Tables 9, 11
Dimension Reduction	2D projection [74, 76, 79, 96]	Prone to pose distortion [76]	Direct extraction from 3D
Computation Complexity	Multi-layer complex processing [96]	Limits real-time application	Light computation Table 10
Matching	ICP Registration [116, 83, 72]	Expensive [83]	Not needed, pose invariant
Reference Face	Neutral Face in 3D-FER [97, 79]	Subject dependent	Not needed, subject independent 3D-FER

can be used as generalized features for both 3D-FR and 3D-FER at a significantly faster computation time without compromising the recognition performance. This suggests the generalizability of the proposed framework in four different recognition tasks (3D-FR, posed 3D-FR, 3D-FER, and posed 3D-FER).

The effect of pose distortions on 3D facial data is not well studied. Prabhu et al. study pose invariant face recognition for 2D facial images by reconstructing their 3D faces [70]. However, they report that at 0.1% FAR, the verification rate drops as low as 44.6% even after using a 3D reconstruction approach. The effect of pose in 3D facial data is mostly resolved by employing a computationally expensive registration method such as Iterative Closest Point (ICP). The proposed framework exploits the inherent geometry of the 3D point cloud data to provide pose invariant features at the cost of very low computation time and complexity. The proposed method shows that if the nose tip location is correctly identified in the posed 3D facial data, the recognition performance in 3D-FR or 3D-FER can remain consistent up to a 55^0 pose angle variation.

Table 12 summarizes the advantages of the proposed framework over the literature in terms of processing steps and methods. Apart from the competitive results shown in Tables 9-11, Table 12 illustrates the benefits of the proposed approach in analyzing 3D facial data for a recognition task. In order to perform both FR and FER using a common set of features for frontal and posed data, the proposed method avoids manual selection of features, segmentation, and expensive pose corrections for each sample of 3D facial data. Secondly, a subject independent 3D-FER is obtained without using the neutral face as the reference to extract features related to different expressions [97, 79]. Thirdly, the proposed work avoids computationally expensive point-based representation and registration of 3D facial data. Finally, the features are directly extracted from the 3D facial data unlike other approaches, where 3D facial data are projected on the 2D space for 3D-FER [74, 76, 79].

4.6 Localization of discriminative 3D facial curves

The radial curve-based representation of 3D face and Frenet frame-based geometric features have been considered in this section to localize facial curves characterizing different facial expressions. For this purpose, the BU-3DFE dataset [103] is used. For each of six basic facial expressions, 3D facial shapes of 50 individuals are considered. Therefore, the dataset contains features from a total of 300 3D faces labeled with six facial expressions. This curve-based localization will be later used in studying several facial muscle activations in individuals with ASD.

4.6.1 Data processing

Following the algorithm proposed in Fig. 11, each radial curve in a 3D point cloud representation of the face is represented by a feature point. Therefore, the weight or contribution of a feature point in classification or discrimination can, in turn, reveal the contribution of a facial curve. Since facial expressions involve multiple facial muscle activations, the aim is to find the most contributing facial curves in rendering such expressions. In this case, the mean curvature (F_κ) of each curve as previously computed in Eq. 21 (Section 4.2.6) is considered.

In this case, two types of localization tasks are performed. In the first case, the supervised feature selection method MRMR is used to rank the features and select the top M number of features from the rank, which best classify the six facial expressions as discussed in Section 4.3.2. The facial curves corresponding to the selected features are identified to illustrate the most discriminative facial regions for classifying six facial expressions. In

(a)   (b)   (c)
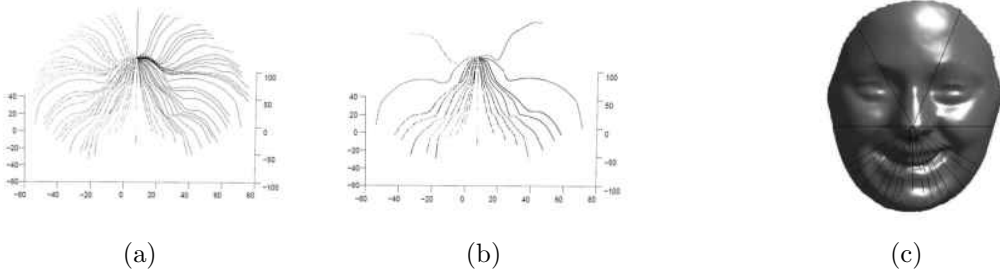
FIG. 19: (a) All 64 radial curves originating from nose tip at the center, (b) Curves corresponding to 40 selected features, (c) Curves following feature selection results are mapped on the 3D face.

the second case, the representative curves for each facial expression are localized. The F_κ feature is obtained for a total of 300 3D faces labeled with six facial expressions as mentioned in Section 4.6. For each individual subject in the BU-3DFE dataset, the squared difference between $F_\kappa(s)$ feature vectors from 3D faces is measured with a particular facial expression and without any expression (neutral). For each facial expression, the squared differences are averaged over 50 individuals to obtain mean squared differences (MSD) as:

$$MSD \;\; = \;\; \frac{1}{50}\sum_{r=1}^{50}\left[F_{\kappa N}(r) - F_{\kappa E}(r)\right]^2, \tag{28}$$

where, $F_{\kappa N}$ and $F_{\kappa E}$ are feature vectors from 3D faces of with neutral and with a given facial expression for the same subject (indexed by r), respectively. The feature points are ranked based on their MDS values, and the top 20 feature points corresponding to the highest MDS values are considered for illustration. Each of these feature points belongs to a radial curve that we can identify and localize on the 3D face.

4.6.2 Results

The MRMR-based multiclass feature selection method shows that 40 out of 64 curves yield the best 3D-FER performance. The optimal number of features or curves is searched from the classification accuracies followed by 10-fold cross validation in classifying six facial expressions.

Considering F_κ feature, a total of 64 feature points are obtained from an equal 64 number of facial curves of a 3D face as shown in Fig. 19(a). The best contributing curves
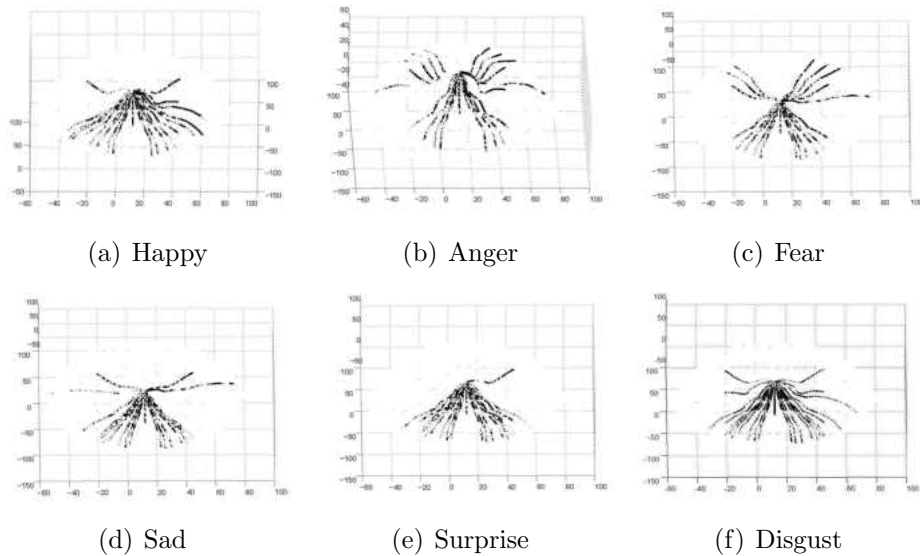
FIG. 20: The 20 most active facial curves on 3D faces for six facial expressions in contrast to the neutral expression.

following the feature selection procedure are shown in Fig. 19(b). These selected curves are mapped back on the facial surface to show that most of the curves lie in the lower half of the face as shown in Fig. 19(c). On the other hand, Figure 20 demonstrates the most discriminative curves contributing to each of six facial expressions following the highest MDS values of the features. Fig. 20 shows that most of the facial expressions involve activations of facial curves in the lower half of the face except for anger and fear expressions. These two facial expressions involve considerable activations of facial curves in the upper half of the face.

The proposed curve-based representation and feature extraction allow classification, recognition of 3D face as well as localization of facial curves based on their activation due to facial expressions. This enables a convenient comparison of 3D faces using the curve based geometric feature representation. Instead of using the entire 3D volumetric surface or 3D point cloud, the curve is easy to acquire, index, and localize on 3D faces for visualization and comparison of the large volume of 3D face. This approach can be employed in the automatic selection of useful regions of the face in a given recognition task. Many studies related to 3D-FR or registration invariably exclude the lower half of the face for analysis [84], [119]. Instead of excluding upper or lower part, one can automatically

select a set of facial curves that perform the best for a given classification or recognition task.

The proposed computational model in this chapter is investigated further in the next chapter to quantify differential traits from the 3D facial data of the subjects with ASD. The geometric feature computed from several facial curves are considered to investigate facial muscle actions local to those curves.

CHAPTER 5

FACIAL MUSCLE ACTIVATION ANALYSIS IN

ASD

5.1 Chapter Overview

The first experiment collects static 3D facial data and video data of the face in response to the visual stimuli and tasks outlined in Chapter 3. This chapter investigates facial muscle-specific response using the static 3D facial data and video data obtained from this experiment. The computational model for analyzing 3D facial data, as discussed in Chapter 4, is utilized to probe facial muscle-specific actions in a non-intrusive manner. Since the proposed geometric features obtained using Algorithm in Fig. 11 are found effective in the classification and localization of facial expressions, these features are used to compute activations of distinct facial curves. The activation of facial curves, in turn, provides the activation the information about activation of the facial muscles local to those facial curves. The investigation with static 3D facial data is complemented with the analysis of video data of the facial responses. A state-of-the-art facial landmark detection and tracking algorithm is used to compute facial muscle-specific activations from the video data. The goal of this chapter is to study potential differential traits in the activation of facial muscles from the geometry of the 3D facial curves and facial landmarks tracked in video data.

5.2 Data Analysis Framework

This section discusses the data analysis frameworks for 2D and 3D facial images, which are used to compute facial muscle movements of the two groups of subjects in the first experiment.

5.3 Analysis of 2D Facial Image

The video camera involved in the first experiment (See Chapter 3) collects facial images of the participants in response to the task and visual stimuli. The video data are sampled

| (a) | (b) | (c) | (d) |
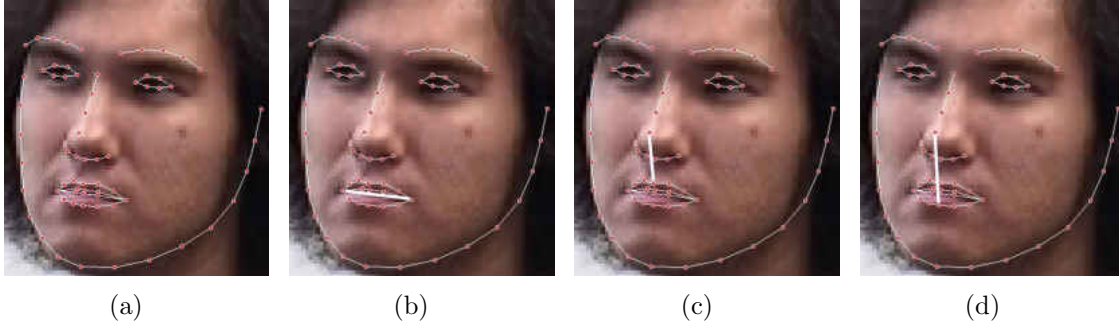
FIG. 21: Illustration of automatic landmark point detection on a 2D facial image frame. (a) 66 landmark points, (b) A line connecting two lip corners, (c) A line connecting the nose tip and the midpoint of the upper lip, and (d) A line connecting the nose tip and the midpoint of the lower lip.

into 2D facial image frames at a rate of one frame per second. This enables the frame-by-frame analysis of facial landmarks movements during the study. A robust facial landmark detection routine is applied, which consistently detects 66 landmark points on each of the 2D facial image frames [65] as shown in Fig. 21(a). Among the 66 landmark points, 12 points trace the edges of the lips, whereas the other points trace the regions of the eyes, nose, and the facial boundary. This automatic landmark-based feature extraction alleviates the need for a registration step, which can be challenging for a 2D facial image with potentially an out-of-plane rotation of the face. Chapter 4 showed that the lower half of the face is the most active part for the rendering of facial expressions (See Fig. 19(b)). The lower half of the face includes the mouth, which contributes to a significant number of facial action units coded by the facial action coding system (FACS) [89]. Therefore, the frame-by-frame movements of the landmark points associated with the lip region have been computed in the following sections. Based on the landmark points of the lip regions, we derive the following features related to the lip action.

(a) Lip corner distance The two lip corners are identified by two distinct landmark points. Figure 21(b) shows a connecting line representing the Euclidean distance between the two lip corners F_{LC} is given as,

$$F_{LC} = \sqrt{(x_{llc} - x_{rlc})^2 - (y_{llc} - y_{rlc})^2}. \tag{29}$$

where, (x_{llc}, y_{llc}) and (x_{rlc}, y_{rlc}) are the coordinate points of left and right lip corners, respectively. This distance feature F_{LC} computes the breadth or spread of the lip corners. The lip corners may be actuated by one or several underlying facial muscles, which, in turn, may change the lip corner distance or lip stretch.

(b) Lip distance The upper and lower lips are primarily actuated by *Orbicular Oris* muscle. This action may manifest as distinct facial actions such as lip funnel, tight lips, pressed lips, lips apart annotated by facial action units 22, 23, 24, and 25, respectively. The nose tip is a rigid or fixed point which can be used as the reference point to compute the relative motion of an individual lip. Following the detection of facial landmarks, we compute the Euclidean distance F_{nul} between the midpoint of the upper lip (x_{ul}, y_{ul}) and the nose tip (x_n, y_n) as follows,

$$F_{nul} \;=\; \sqrt{(x_n - x_{ul})^2 - (y_n - y_{ul})^2}. \tag{30}$$

Figure 21 (c) shows a line as the distance between the detected midpoint of the upper lip and the nose tip. Similarly, Figure 21(d) illustrates a line as the distance between the midpoint of the lower lip (x_{ll}, y_{ll}) and the nose tip (x_n, y_n), which can be calculated as follows,

$$F_{nll} \;=\; \sqrt{(x_n - x_{ll})^2 - (y_n - y_{ll})^2}. \tag{31}$$

(c) Analysis of the features The proposed distance features depend on the participant's facial morphology and may vary from participant to participant. A comparable physiological response, independent of the participant, can be obtained from the mean deviation of the proposed feature over a set of facial image frames. In this regard, a pre-stimulus facial image frame with neutral expression is considered as the reference facial image frame. In each trial, the facial image frame at 1000 ms after the onset of the stimulus is compared with the pre-stimulus reference facial image frame. For each participant, the mean deviation (Q) is computed by averaging the absolute differences between the pre-stimulus features (F_N) from the reference facial frame and N features (F_i) from N facial frames in response to the visual stimuli as,

$$Q = \frac{1}{N} \sum_{i=1}^{N} abs(F_N - F_i). \tag{32}$$

The mean deviation of the facial muscle action for each participant is considered to perform within-group and between-group ANOVA tests. The ANOVA tests will reveal if there is any significant difference in the mean deviation of the facial muscles within a group or between two groups.

5.4 Analysis of 3D Facial Data

The 3D optical camera yields a dense 3D point cloud representation of the face. The 3D facial point cloud data can be compared in terms of surface-based representation encompassing the entire face, which may be computationally expensive and prone to errors due to a registration step. Instead of considering the entire 3D facial point cloud, a small set of 3D curves related to different facial muscle regions are automatically extracted for analysis. The mean curvature feature F_κ, derived in Chapter 4, is computed from the geometry of the extracted 3D facial curve. The mean curvature feature will capture the change in the facial muscle movement and can serve as a metric to compute the physiology of the facial muscle. The following sub-sections discuss the preprocessing and feature extraction steps for the 3D facial point cloud data.

(a) **Preprocessing and Normalization** The 3D optical camera system reconstructs a 3D facial point cloud from the captured stereoscopic images, which include undesired regions protruding from the facial surface such as ears, scalp, and partial neck. There are a few floating 3D points around the surface of the face. Publicly available software application named *MeshLab* (http://meshlab.sourceforge.net/) facilitates the visualization and manipulation of the 3D facial point cloud data. This software application is utilized to remove the undesired floating points and protruding regions to obtain the useful region of the face from the 3D point cloud. The 3D facial point is then represented by depth values along the z-axis, where z-values are essentially mapped on to non-uniform sample points (x, y) in Cartesian coordinate as z=f(x, y). In order to have a common topology for feature extraction, the z-values are interpolated on a uniformly sampled square grid (x', y'), which is defined within the boundary values defined in Eqs. 1 and 2 (See Section 4.2.2). This interpolation step normalizes all the 3D facial point clouds on a common topology of a uniformly sampled grid. Following the normalization, an Iterative Closest Point (ICP) based rigid registration is applied to align a test 3D facial point cloud with the reference 3D facial point cloud.

(b) Extraction of facial curves Following the method discussed in Chapter 4, radial curves are extracted from 3D facial data corresponding to five distinct locations of the face. Following the previously described method, the nose tip of the 3D facial point cloud is detected as the highest z value and the coordinate point of the nose tip projected on the uniform grid (x_n, y_n) is used to draw a linear path defined by a slope ratio as follows.

$$\frac{p}{q} = \frac{y_n - y_u}{x_n - x_u},\tag{33}$$

where (x_u, y_u) are the points of the uniform grid that satisfy the equation of the linear paths drawn as follows.

$$y - y_n = \frac{y_n - y_u}{x_n - x_u}(x - x_n).\tag{34}$$

The z_u values corresponding to (x_u, y_u) points belong to a 3D curve on the 3D facial surface, which can be represented in an angular form as θ_{ij} as follows,

$$\theta_{ij} = \frac{k_i \pi}{2} + \arctan\left(\frac{p}{q}\right)_j, k = 1, 2, 3 \text{ and } 4,\tag{35}$$

where the uniform grid is divided into four quadrants denoted by four different values of k, and $\frac{p}{q}$ stands for the slope ratio corresponding to a linear path j in a given quadrant. This representation can be used to localize and index infinite number of radial curves around the face starting from the nose tip.

Taking advantage of this mathematical representation, five radial curves are extracted corresponding to the five facial muscle regions as shown in Fig. 22. Although each curve originates from the nose tip, the curve segment lying on the nose region is discarded, and only the curve segment corresponding to the facial surface is considered for the analysis of facial muscle activations. The five probe muscle regions are related to: *Zygomaticus Major* Right (ZMR) and Left (ZML), *Levator Anguli Oris* Right (LAOR) and Left (LAOL), and *Orbicularis Oris* (OrOr). Following Eq. 37, Table 13 shows the parameter values for extracting the five radial curves corresponding to the five probe facial muscle regions. In Table 13, $\frac{p}{q}$ represents the slope ratio of the linear path that extracts the facial curve in one of four quadrants (indexed by k), whereas θ represents the angular orientation of the facial curve originating from the nose tip.
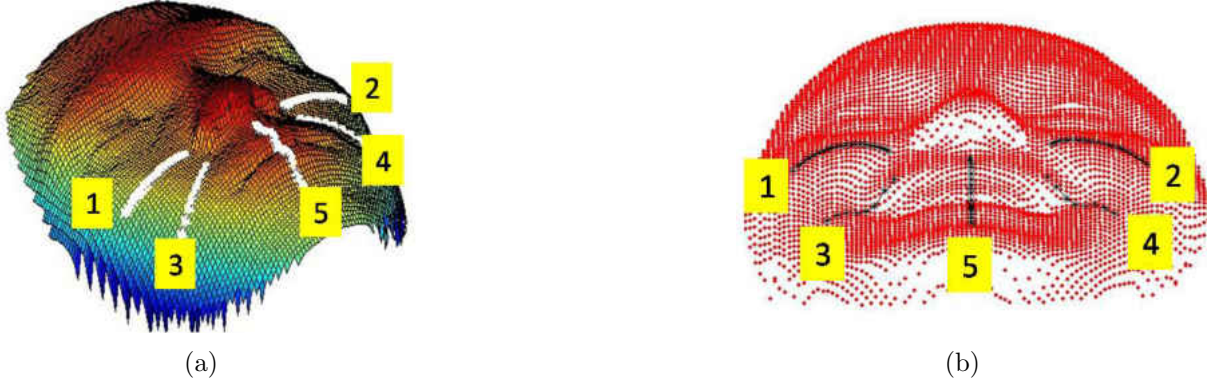
FIG. 22: Five 3D facial curves and their locations. 1= *Zygomaticus Major* Right (ZMR), 2=*Zygomaticus Major* Left (ZML), 3=*Orbicularis Oris* (OrOr) ,4= *Levator Anguli Oris* Right (LAOR), and 5=*Levator Anguli Oris* Left (LAOL). (a) Five facial curves on a mesh representation of a 3D face and (b) on a topological view of a 3D facial point cloud.

(c) Curve-based geometric features The extracted 3D curves are further processed to obtain geometric features that represent the change in the geometry of the curve due to the activation of local facial muscle. The mean curvature of a 3D space curve can be computed by averaging all the curvature values of the constituting N number of 3D points as follows,

$$F_\kappa = \frac{1}{N} \sum_{i=1}^{N} \kappa(s_i). \tag{36}$$

The curvature represents how the curve bends along its path at a particular point. The mean curvature of a 3D curve thus characterizes the average bending of the entire curve along its path. The results in Chapter 4 shows the effectiveness of this feature in the classification and localization of facial responses for six facial expressions (See Section 4.6.2). These benefits of the proposed feature can be utilized to analyze facial muscle-specific responses from 3D faces. A 3D face with neutral expression is compared with a 3D face in response to a visual stimulus (test 3D face) in terms of their respective mean curvatures obtained from the 3D curves. To facilitate this comparison, the absolute difference between the mean curvature feature values from a test 3D face and a neutral 3D face of the same participant is computed. The absolute differences are then averaged over

TABLE 13: Parametric values to extract 3D radial curves related to different facial muscle regions.

Facial Muscle Regions	Slope Ratio $(\frac{p}{q})$	Quadrant Index (k)	Angular Orientation (θ)
Zygomaticus Major Right (ZMR)	$\frac{1}{0}$	1	180^0
Zygomaticus Major Left (ZML)	$\frac{-1}{0}$	1	0^0
Levator Anguli Oris Right (LAOR)	$\frac{2}{1}$	2	206^0
Levator Anguli Oris Left (LAOL)	$\frac{-2}{1}$	4	333^0
Orbicularis Oris (OrOr)	$\frac{0}{1}$	3	270^0

the number (M) of available test faces as follows,

$$\overline{\Delta F_\kappa}(j) = \frac{1}{M}\sum_{i=1}^{M}|F_{test}(i,j) - F_{neutral}(j)|, \tag{37}$$

where, $\overline{\Delta F_\kappa}$ is the mean change in curvature (MCC) of the facial curve, which is utilized as a feature to represent facial muscle actions of an individual participant, and j is the index for the facial curves ($j =1,2,..5$). This feature representation includes the average of absolute differences between two facial samples, which can cancel out potential errors due to reprocessing of 3D facial data.

Similar to the 2D facial image analysis, ANOVA tests are performed on the MCCs of the facial curves corresponding to different facial muscles. A within-group ANOVA test is conducted to find any significant differences among five different muscle actions within a particular group. To measure the statistical difference between MCCs of each pair of facial muscles, a post-hoc Tucky test is conducted where the degrees-of-freedom of F-statistics are given as F(r, t). Here, r and t are computed as (2-1 =1) for a pair of muscles and (8-1)X 2 =14 for eight subjects, respectively. The significance or α level is chosen as 0.05 for this study.

5.5 Results

This section illustrates the results following the processing steps and methods presented in the previous section.

5.5.1 Results for 2D Facial Images

For each participant, the mean deviation of a lip action over 12 facial image frames in contrast to a pre-stimulus neutral facial image frame is computed. These 12 facial images are taken from 12 trials of the study, where each frame is shot at 1,000 ms after the onset of the visual stimulus in the trial. For each of the 12 trials, the average of mean deviation of lip actions is taken over all participants in a group to represent a group characteristic.
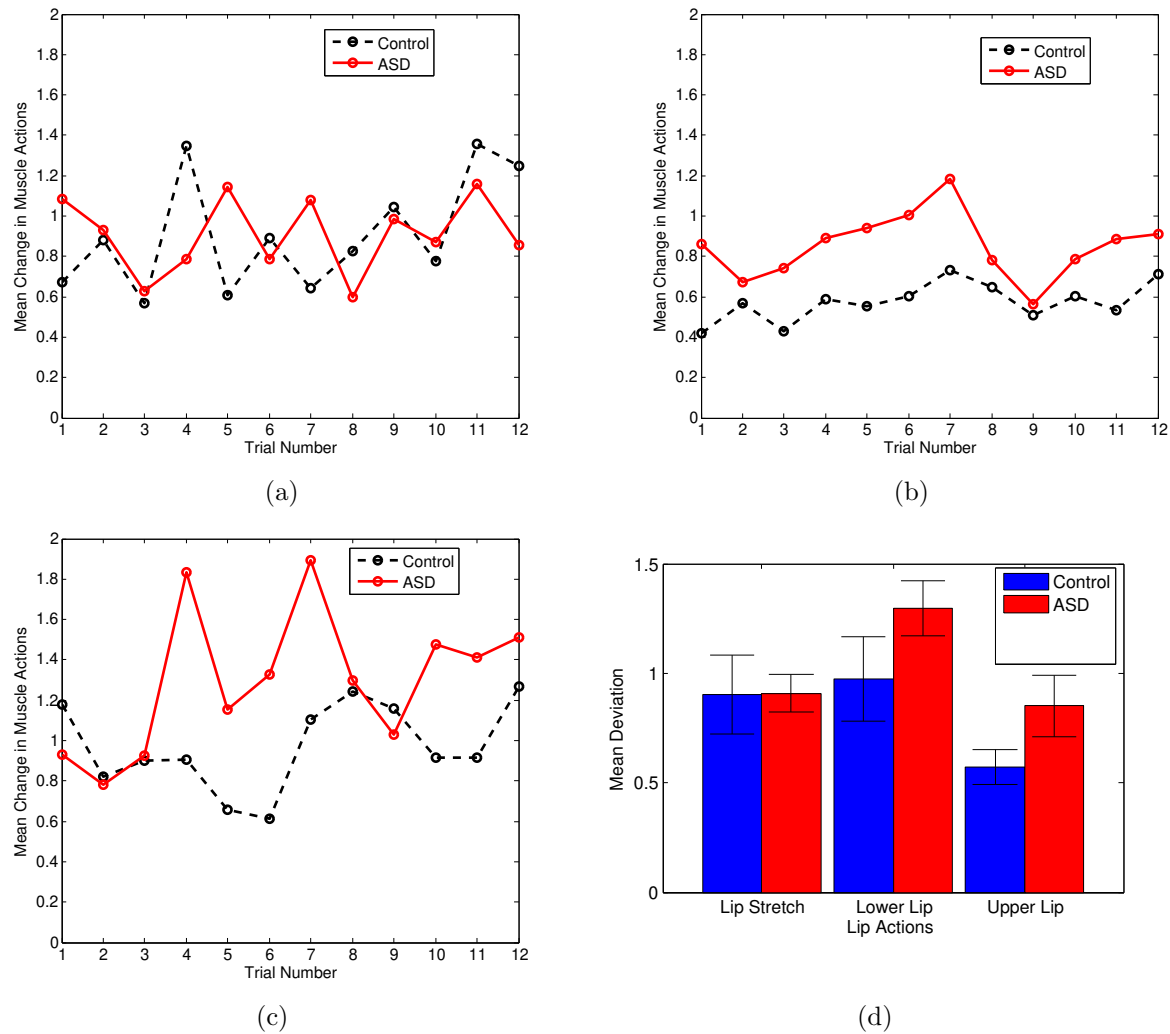
FIG. 23: Illustration of lip actions for the two groups of participants. The mean change in (a) Lip corner distance, (b) Upper lip action, (c) Lower lip action. (d) Mean and standard error plots for different lip actions over all 12 trials.

TABLE 14: ANOVA tests for different lip actions in the ASD and TD groups. The upper and lower triangles for within-group column show the results for the TD and the ASD group, respectively. The significant p values are highlighted.

Lip Actions	Within-group			Between-group
	Lip Corner Distance	Lower Lip	Upper Lip	
Lip Corner Distance	—	$F(1,14)=0.06,$ $p>0.05$	$F(1,14)=3.57,$ $p>0.05$	$F(1,14)=0.21,$ $p>0.05$
Lower Lip	$F(1,14)=6.67,$ **p<0.05**	—	$F(1,14)=4.11,$ $p>0.05$	$F(1,14)=3.13,$ $p>0.05$
Upper Lip	$F(1,14)=0.5,$ $p>0.05$	$F(1,14)=5.74,$ **p<0.05**	—	$F(1,14)=2.5,$ $p>0.05$

Fig. 23(a) shows the mean change in the lip corner distance or lip stretch at different trials for the ASD and TD groups. Both groups show a similar trend of lip stretch actions. Taking the mean over all 12 trials, Fig. 23(d) shows no obvious difference between the two groups in terms of lip corner distance or lip stretch. The ASD group manifests a higher action in the upper lips when compared with the TD group as illustrated in Fig. 23(b). The average from all 12 trials reveals a higher mean deviation of the upper lip for the ASD group (See Fig. 23(d)). Similar observations are found for the lower lip action, which is evident in Figs. 23(c) and (d). In general, Fig. 23(d) reveals that the ASD group manifests a higher mean deviation in all lip actions when compared with the TD group.

A within-group ANOVA test reveals no significant difference among different lip actions within the TD group, $F(2,21)=1.18$, p>0.05. However, for the ASD group, there remains a significant difference among the three different lip actions, $F(2,21)=4.14$, p<0.05. To find further effect on the pair-wise lip actions, we conduct a post-hoc Tucky test within each group. The upper triangle of the within-group column in Table 14 shows the pair-wise statistical results for the TD group. No lip action pair is found to be statistically different for the TD group. The lower triangle of the within-group column in Table 14 shows the statistics for the ASD group. For the ASD group, there exists a significant difference between the lip stretch and lower lip actions, $F(1, 14)=6.67$, p<0.05, as well as between the upper lip and lower lip actions, $F(1, 14)=5.74$, p<0.05.

TABLE 15: Post-hoc Tucky tests on mean change in curvature (MCC) of different 3D facial curves representing different facial muscles. The upper and lower triangles show the results from the control and ASD groups, respectively. The significant p values are highlighted.

	ZMR	ZML	OrOr	LAOR	LAOL
ZMR	—	F=0.39, $p > 0.05$	F=117, **p<0.0001**	F=2.28, $p > 0.05$	F= 4.60, **p<0.05**
ZML	F=0.47, $p > 0.05$	—	F=130, **p<0.0001**	F=5.24, **p<0.05**	F=7.75 , **p<0.05**
OrOr	F=12.51, **p<0.01**	F=22.68, **p<0.001**	—	F=81.65, **p<0.001**	F=83.22, **p<0.001**
LAOR	F=0.06, $p > 0.05$	F=1.34, $p > 0.05$	F=18.7, **p<0.001**	—	F = 0.13, $p > 0.05$
LAOL	F=1.52, $p > 0.05$	F=0.21, $p > 0.05$	F=48.81, **p<0.0001**	F=5.4, **p<0.05**	—

5.5.2 Results for 3D facial images

For a participant, the mean change in curvature (MCC) is computed for each facial muscle region over all the available 3D facial point cloud data. For the TD group, a within-group ANOVA test reveals a statistically significant difference among the MCCs from five muscle regions, $F(4, 35) = 47.49$, $p < 0.001$. Out of five probe muscles regions, four of them lie symmetrically at the left and right sides of the face. These muscle regions are ZMR and ZML as well as LAOR and LAOL. A post-hoc Tucky test shows further inferences about the pair-wise muscle actions in the upper triangle of Table 15 for the TD group. The upper triangle of Table 15 shows that ZMR and ZML muscles have no significant differences in activations, $F(1,14) = 0.39$, $p > 0.05$, which is also true for LAOR and LAOL muscles, $F(1, 14) = 0.13$, $p > 0.05$. These results indicate a symmetric action in the facial muscles of the TD group. The upper triangle of Table 15 also demonstrates that the *Zygomaticus Major* muscle yields significantly different MCC than that of the *Levator Anguli Oris* muscle for the TD group. This difference is evident between ZMR and LAOL muscles as $F(1, 14)=$
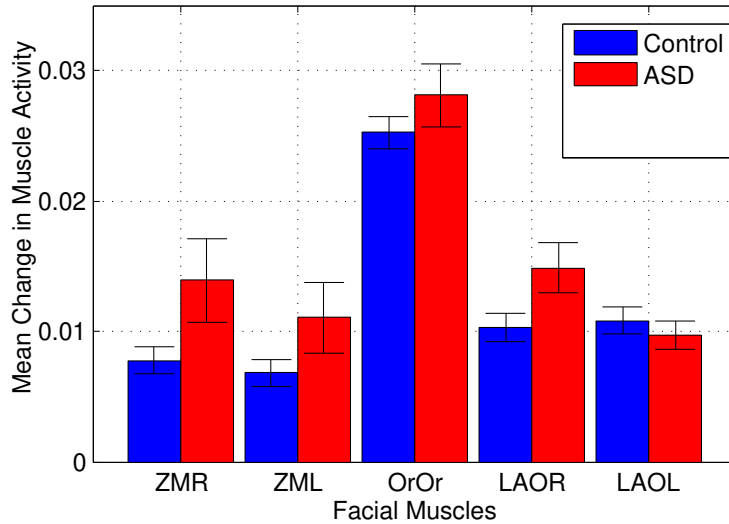
FIG. 24: Mean change in curvature (MCC) for five different probe muscle actions. *Zygomaticus Major* Right (ZMR), *Zygomaticus Major* Left (ZML), *Orbicularis Oris* (OrOr), *Levator Anguli Oris* Right (LAOR), and *Levator Anguli Oris* Left (LAOL).

4.60, p <0.05, between ZML and LAOL muscles as $F(1, 14)$=7.75 , p <0.05, and between ZML and LAOR as $F(1, 14)$=5.24, p <0.05. The OrOr muscle responsible for lip actions is found to be the most active facial muscle yielding the highest MCC as shown in Fig. 24.

Similar to the observations in Table 15, Fig. 24 illustrates the symmetry in the facial muscle actions for the TD group in the ZMR-ZML and LAOR-LAOL muscle pairs. Similar to the TD group, ANOVA test reveals significantly different MCC for the five facial muscle regions within the group diagnosed with ASD, $F(4, 35)$ =9.57, p <0.001 The lower triangle of Table 15 demonstrates the post-hoc Tucky test results for each pair of the facial muscles in the ASD group. The symmetrically located facial muscles (ZML-ZMR) do not yield any significant difference in between, $F(1, 14) = 0.47$, p >0.05. However, unlike the TD group, the other symmetrically located muscle pair LAOL-LAOR yields significantly different MCC, $F(1, 14)$ =5.4, p <0.05. This indicates an asymmetry in the facial physiology of the ASD group.

The subject-specific plots of left and right facial muscle actions for the two groups are presented in Fig. 25. Fig. 25(b) shows that the average facial responses of three subjects, out of eight, demonstrate left and right asymmetry in their facial muscle activations. Two of them are siblings of 11 and 13 years of age who are expected to have similar treatment
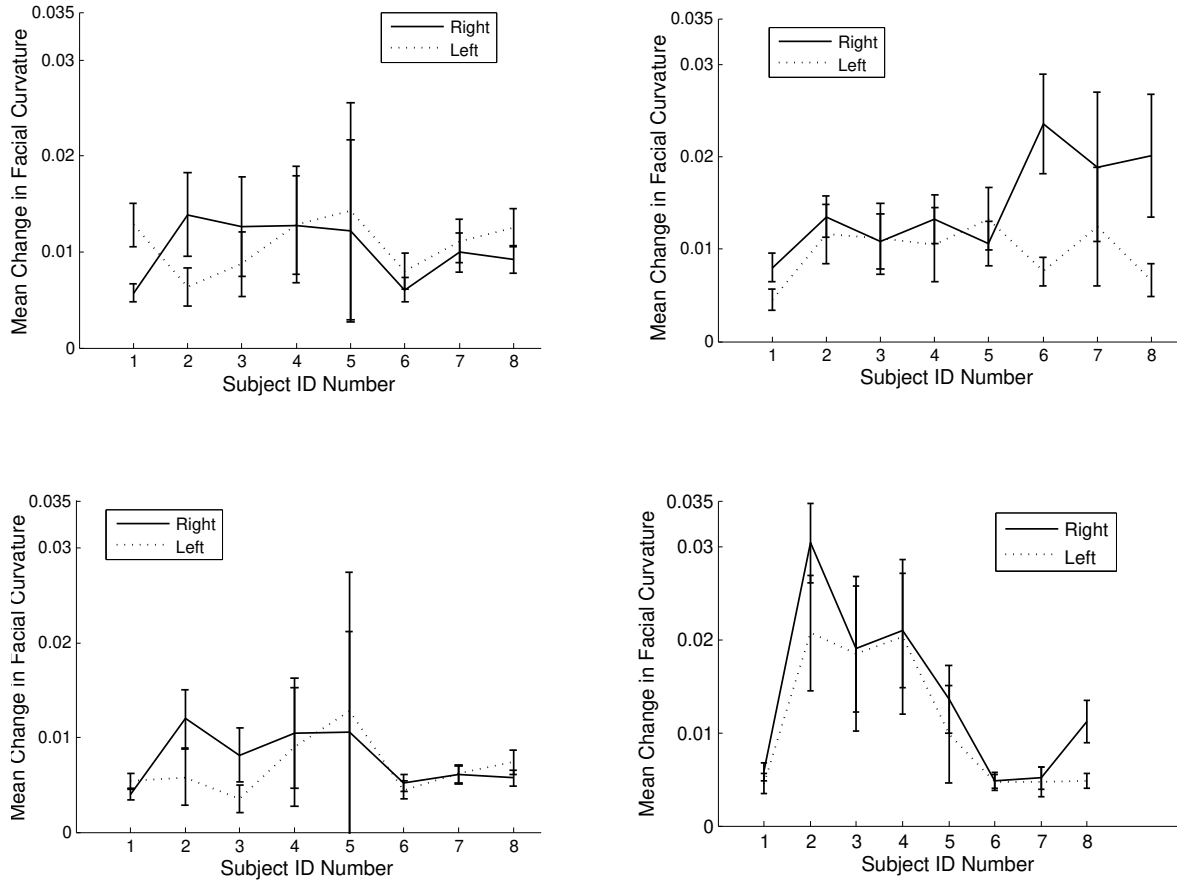
FIG. 25: Subject-specific mean change of the facial curvature feature in left and right: for *Lavator anguli oris* muscle regions of (a) Control group and (b) Group with ASD, for *Zygomaticus major* muscle regions of (c) Control group and (d) Group with ASD.

history. The visual inspection of their video images also reveals exaggerated lip actions. The other individual is a 17-year old. The 17-year old consistently shows smiles most likely because he had been reinforced so much to respond with a smile. Furthermore, the attempt to cope with anxiety and/or stress can stimulate similar repetitive motor actions in individuals with ASD.

Unlike the TD group, the lower triangle of Table 15 also reveals that the group with ASD fails to yield significantly different muscle actions between the *Zygomaticus Major* and *Levator Anguli Oris* muscles. This finding is evident between ZMR and LAOR as $F(1, 14)= 0.06$, $p > 0.05$, between ZMR and LAOL as $F(1, 14)=1.52$, $p > 0.05$, between ZML and LAOR as $F(1, 14)=1.43$, $p > 0.05$, and between ZML and LAOL $F(1, 14)=0.21$,

p >0.05. Similar to the TD group, the OrOr muscle yields significantly higher MCC when compared with other muscles within the ASD group. A between-group design ANOVA is performed to investigate any significant difference between the two groups in terms of the MCC for a particular facial muscle. Figure 24 shows a comparative plot of MCCs produced by the five facial muscles for the two groups. Fig. 24 shows that the ASD group yields higher MCC than the TD group for all facial muscles except for LAOL. However, none of the facial muscle regions shows significant difference between two groups. The between-group statistics for the five facial muscle regions are as follows: for ZMR is $F(1, 14)$=3.29. p=0.09, for ZML is $F(1, 14)$=2.16, p=0.163, for LAOR is $F(1, 14)$=4.19, p=0.059, for LAOL is $F(1, 14)$=0.58, p=0.45, and for OrOr is $F(1, 14)$= 1.13, p=0.30.

5.6 Discussions of the Results

The experiment using 3D facial images was performed on eight subjects with ASD. Each 3D facial image is captured only once after 1000 ms of the onset of the stimulus in each trial. This yields twelve 3D facial images per subjects, which are used to compute the average facial response for each subject. The average facial responses for these three subjects contributed to a significant inference for the group of eight subjects. These subject-specific facial responses may have impacted the group statistics since the group consists of a limited number of subjects. While a larger dataset would yield more reliable group statistics, the subject-specific cases can provide us with useful differential traits for the individual facial responses.

This study demonstrates the feasibility of the proposed non-intrusive computational method in the analysis of the physiology of facial muscles. The proposed methods show that a particular facial muscle region can be non-intrusively probed from 2D facial image frames or 3D facial point cloud data to analyze its physiology. These methods are tested in a pilot study to investigate the atypical facial muscle actions for the group with ASD in contrast to the TD group. In general, the between-group ANOVA test reveals no significant difference between the ASD and TD groups in terms of a facial muscle action. However, a borderline difference appears between the two groups in the case of LAOR muscle action using 3D facial point cloud data, $F(1, 14)$=4.19, p=0.059. This inference may turn significant with a larger size of population and 3D facial point cloud samples. Furthermore, within-group ANOVA tests reveal several interesting inferences about the facial muscle actions as follows.

Intense facial expressions in participants with ASD   Both of the facial imaging modalities (2D and 3D) confirm a higher magnitude of facial muscle actions in the ASD group when compared with the TD group. This is evident in Fig. 23(d) and Fig. 24 for 2D and 3D facial images, respectively. This observation may provide quantitative support for similar observations found in the prior studies [16, 53]. In an EMG-based physiological study, Mathersul et al. report higher *Zygomaticus Major* muscle response in the ASD group when compared with the TD group [53]. In a study involving visual evaluation of facial image frames, an intense level of facial expression is reported in the ASD group when compared with the TD group [16]. However, prior subjective studies report that such intense facial expressions in the group with ASD has lack of natural traits, which is reported as the oddity in facial expressions [16] or awkward expression [17]. Consequently, this dissertation provides further quantitative supports to discover the differential traits in the facial expressions of the group with ASD.

Asymmetry in the facial expressions of participants with ASD   The oddity in facial expressions may be explained by the asymmetry in the physiology of facial muscles as observed in this study. From the analysis of 2D facial image frames, the ASD group reveals significant differences between the mean deviations of the lower and upper lips. This difference is found insignificant in the TD group. On the other hand, the analysis of 3D facial point cloud data reveals significant difference between the MCCs of left and right *Levator Anguli Oris* muscle regions for the group with ASD. All these observations indicate asymmetric facial muscle actions in the ASD group, and such asymmetry is not statistically significant for the TD group. A few recent studies report a significant structural asymmetry between the left and right halves of the face in participants with ASD when compared with the TD group [34, 36]. Apart from the morphological facial asymmetry, this study shows asymmetry in the physiology of facial muscles, which may contribute to the manifestation of oddity in the facial expressions.

Lack of differential facial muscle actions in participants with ASD   The statistical analysis also reveals that the TD group yields a significant difference in actions between the *Levator anguli oris* (LAOL, LAOR) and *Zygomaticus major* (ZML and ZMR) muscle regions. This indicates that these two muscle regions are activated independently for the TD group. Although the participants with ASD exhibit a higher mean change in facial

muscle actions when compared with the TD group, the ASD group fails to produce significant differences in these two specific muscle actions. This may indicate the inability of the participants with ASD to control these muscles independently. Such muscle actions may blend together, which may potentially contribute to the manifestation of oddity in facial expressions.

This Chapter investigates the proposed curve-based computational model for the detection of oddity in the facial expressions of individuals with ASD. Statistical analyses are performed to study the physiology of facial muscles in a non-intrusive manner using the geometric feature of 3D facial curves. Statistical analyses reveal intense, but asymmetric facial muscle actions in the subjects with ASD, which may objectively define measurable impairments in exhibiting facial expressions. Such facial oddities may be one of the differential traits for the group with ASD.

CHAPTER 6

BEHAVIORAL MARKER DETECTION IN ASD

6.1 Chapter Overview

This chapter investigates differential traits of ASD from facial expressions and eye-gaze data using computer vision-based methods. The nonverbal items of the BPASS protocol such as eye contact, facial expressions, and social smile, as discussed in Section 1.2, are studied using eye-tracker and both video and static 3D facial image data, respectively. For the time first time in ASD literature, FACS is employed to encode and classify FAUs from facial image data as mentioned in Chapter 2. The goal of this chapter is to identify a behavioral marker from three categories of response: prevalence of FAUs in facial images, subjects perception of the displayed facial expressions, and eye-gaze fixation while the subjects perform the same tasks assigned in the first experiment. The behavioral markers gleaned from the physiology of the face and vision may further facilitate the computation of the severity and prognosis of the disorder. The findings related to the behavioral marker have implications in the early detection of ASD-related traits to devise an effective intervention strategy.

6.2 Computer-based Analysis of Facial Response

This section presents the proposed computational methods for computer-based recognition of FAUs from 2D and 3D facial images.

6.2.1 Dataset of Facial Action Units

For the computer vision-based classification of FAU, we use a publicly available dataset of 2D and 3D facial images benchmarked with different FAUs. The dataset is known as the Bosphorous facial expression dataset [111]. On an average, the dataset provides 2D and 3D facial images of fifty individuals annotated by professional FACS coders. In this study, five FAUs, AU 10, AU 12, AU 14, AU 24, and AU 25, are probed, as mentioned in Section 2.5.3. The dataset is used to obtain a set of ground truth facial features that

Facial Image

$\downarrow$

Preprocessing

$\downarrow$

Registration & Normalization

$\downarrow$

Lower Face Segmentation

$\downarrow$

Gabor Feature Extraction

$\downarrow$

Feature Selection

$\downarrow$

Final Feature
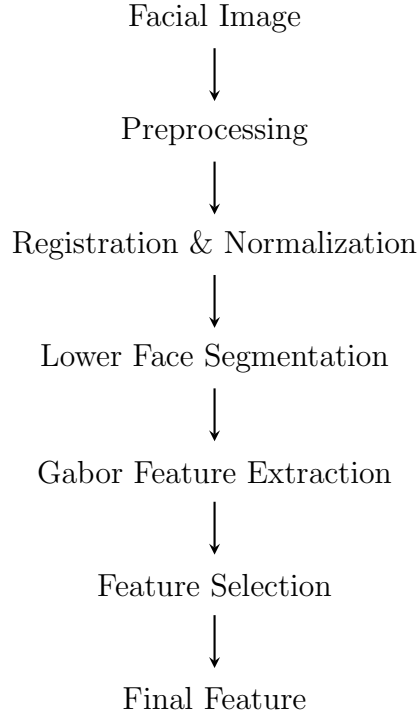
FIG. 26: Steps for preprocessing and feature extraction from facial images prior to the classification of FAUs.

represent a particular FAU. These ground truth features are used to train the classifier models for detecting a FAU. Since typical facial expressions are compound and constitute more than one FAU [92, 91], we design five binary classifier models corresponding to five FAUs. The five binary classifiers are AU 10 Vs Neutral, AU 12 Vs Neutral, AU 14 Vs Neutral, AU 24 Vs Neutral, and AU 25 Vs Neutral. These classifiers are trained by the ground truth features extracted from the Bosphorous dataset. The following sections describe the data processing and feature extraction steps involved in the analysis of 2D and 3D facial images.

6.2.2 Analysis of Facial Data

The basic pipeline for analyzing facial data is illustrated in Fig. 26 and discussed in the following subsections for both 2D and 3D facial data.

Preprocessing: A 15-second video clip corresponding to a trial of the study is sampled into 15 2D-facial image frames at the rate of one frame per second. A robust facial landmark detection algorithm is adapted from the work of Zhu et al. [120], which can detect facial

landmarks on eyes, nose, mouth, and facial edges within a complex color background. The landmarks on facial edges are used to determine a rectangular boundary for the facial region, which is segmented from the background of the raw video frame.

Unlike 2D facial images with intensity values, 3D faces are represented by point clouds in 3D space. A raw 3D face contains non-facial regions such as the neck, hair, partial scalp, ears, and other non-facial outliers which are removed to obtain an effective facial region. An open-source software application known as MeshLab (http://meshlab.sourceforge.net/) is used to visualize and clean each 3D facial point cloud. In addition, the 3D point clouds generated from faces of younger subjects are typically smaller in shape than their older peers. Therefore, the smaller 3D faces are uniformly normalized (scaled up) using MeshLab to an average face of the age group without affecting the local variations in the facial images.

Registration and Normalization: While 3D point cloud data can be conveniently corrected for pose and scale, 2D intensity-based facial images suffer from these non-trivial challenges. The proposed robust facial landmark detection algorithm [120] is used to facilitate a rigid registration of a probe 2D facial image with a reference image. The coordinate points of the detected landmarks corresponding to the probe face and the reference face are considered to compute a rigid transformation matrix, which is applied to the probe image for registration. The registered 2D facial images are resampled and normalized as N-by-N dimension.

For the 3D facial point cloud, an iterative closest point (ICP) based rigid registration is applied to automatically align all 3D facial point clouds to a reference 3D face. Following the registration, a 3D facial point cloud is interpolated on a 200x200 uniformly sampled grid (x', y') within the boundary values defined by Eqs. 1 and 2 (See Section 4.2.2).

The 3D facial surface points z=f(x, y) are linearly interpolated on a grid of uniform sample points as z'=f(x', y'), where z represents the interpolated depth of a 3D point on a grid point (x,y). Maximum surface curvature is computed at each of the grid point for the surface z=f(x,y) as used in a recent work on detecting FAUs from 3D facial data [74].

Feature extraction: The lower region of the normalized N-by-N 2D facial image is segmented for feature extraction. The color images are converted to gray scale as in Fig. 27(a). A bank of 2D Gabor functions are used as Gabor filters with 5 scales and 8 orientations as shown in Fig. 27(b) and the extracted Gabor features are shown in Fig. 27(c). A total of 40 Gabor filters have been constructed to extract Gabor features from the segmented lower face. For the 3D facial point cloud, the maximum surface curvature feature is computed at each point (x', y') of the proposed uniform grid for the surface
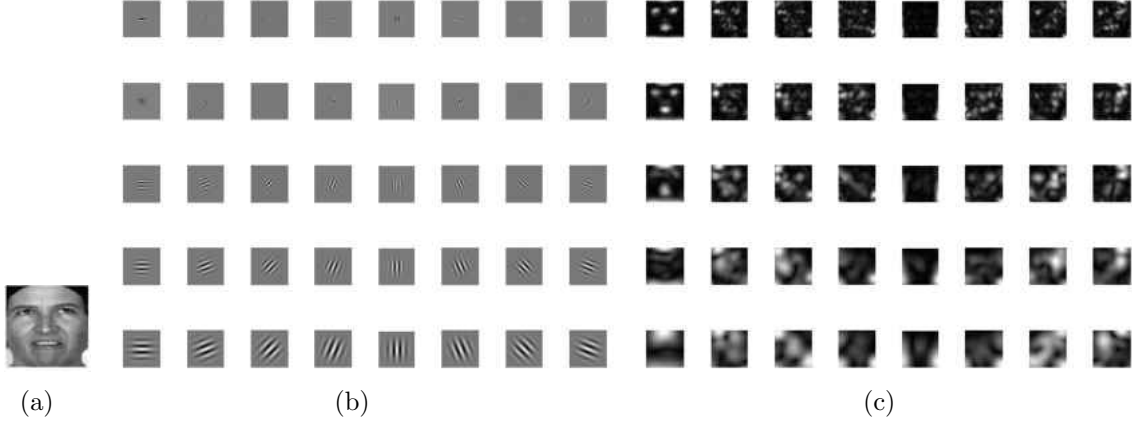
FIG. 27: Gabor filters and filter outputs. (a) Input facial image, (b) 40 Gabor filters, (c) Extracted Gabor features from the input facial image.

z=f(x',y'). The curvature values computed on the points of 3D surface are projected on to the 200x200 grid as gray scale intensity values [74]. Since the target FAUs appear in the lower half of the face, the curvature map of the 3D surface is segmented to account only the lower half of the face for feature extraction. Similar to the step for 2D facial images, 40 Gabor filters are used to extract the Gabor features from the maximum curvature feature map of the 3D face. A similar approach involving the Gabor features is applied in the state-of-the-art studies on FAU recognition [94, 74].

Feature Selection: Once the Gabor features are extracted, the same feature selection technique is applied for both 2D and 3D facial images. This is to reduce the large dimension of Gabor feature into a set of representative features. The Gabor features are ranked based on their contribution in differentiating between a FAU and a neutral expression. The benchmarked 2D and 3D faces from the Bosphorous dataset are employed to identify a subset of features that yield the best performance for the binary classifier. The Gabor features are extracted from the 2D and 3D facial images corresponding to five different FAUs and neutral expressions, respectively. The mean squared differences (MSD) for a FAU over N (=50) 2D or 3D facial image samples is obtained as,

$$MSD = \frac{1}{N} \sum_{k=1}^{N} \left[F_N(:,k) - F_{FAU}(:,k) \right]^2 , \tag{38}$$

where F_N and F_{FAU} represent Gabor feature vectors from the kth subject in the dataset corresponding to neutral expression and a particular FAU, respectively. The indices of the
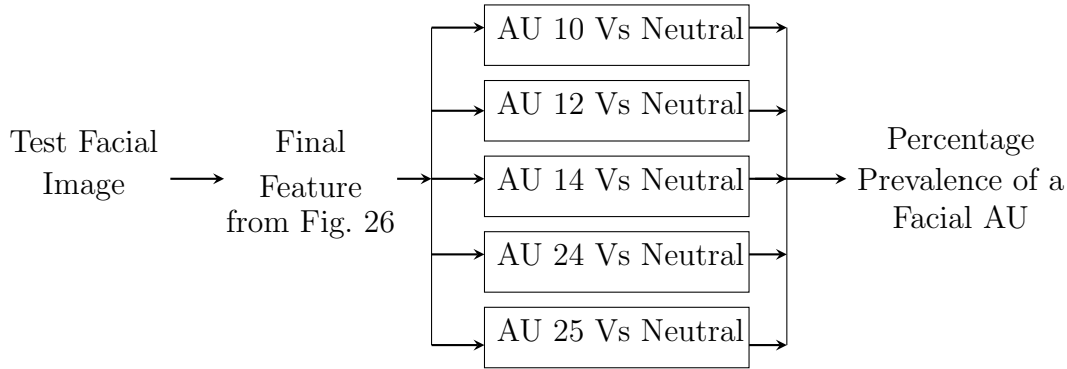
FIG. 28: Classification and evaluation of the test facial images through five binary classifier models corresponding to five target FAUs.

feature vector are ranked based on the MDS values. The indices corresponding to MSD values higher than a threshold are stored for each binary classifier. These indices are used to select the representative features from the large Gabor feature vector of a probe facial image.

Evaluation Methods: The selected Gabor features acquired from the test or training datasets (Fig. 26) are considered for a three-step evaluation process as follows.

In the first step, a ten-fold cross-validation is performed for each binary classifier using the selected features from the training dataset. An average of the classification accuracies from the ten folds represents the performance of a particular binary classifier. This step is to demonstrate the performance of our proposed FAU classification framework, as well as to optimize the classifier parameters to yield the best performance. In the second step, the outcomes and optimization of classifiers in the cross-validation step are taken into account to construct five binary classifier models (See Fig. 28). A binary classifier is trained by the selected features from ground truth facial image samples with neutral expression and a target FAU, respectively. In the third step, the trained binary classifiers are used to detect FAUs from probe 2D and 3D facial data collected from the participants in this study, respectively. A probe facial image may contain more than one FAU as mentioned in Section 6.3.2. For a number of available facial images for a subject, we compute the percentage of prevalence of each FAU from these facial images. For a probe subject, the percentage of prevalence of an action unit X (PAUX) is computed as follows,

$$PAUX = \frac{NAU}{TNFI} \times 100, \tag{39}$$

FIG. 29: Four visual areas of interest (AOI) for eye-tracking. UP and LP represent the upper and lower part of the face, respectively. GUI shows the graphical user interface region. The rest belongs to the *No region.*

where NAU is the number of facial images detected with the AU X, TNFI denotes the total number of facial images available from the probe subject, and X represents any one of the probe FAUs such as 10, 12, 14, 24, or 25.

Percentage of prevalence accounts for the probability of a FAU to appear in a subject's face while participating in the study. The percentage of prevalence of different FAUs from the participating subjects are statistically evaluated in between-group and within-group designs of ANOVA tests. These tests will reveal any significant impact of the experimental task and stimuli on the prevalence of FAUs. A significance level α is equal to 0.05 for the ANOVA tests.

6.2.3 Analysis of Eye-tracker data

The eye-tracker system provides coordinate location and gaze duration for each gaze fixation point. Gaze fixation is initiated at the onset of the visual stimulus until the disappearance of the GUI when the subject chooses an answer. A gaze fixation with equal or more than 320 ms duration is considered as voluntary gaze fixation for this study [121]. Each gaze fixation is labeled by one of four possible visual areas of interest (AOI) based on its coordinate location in the visual scene as shown in Fig. 29. These AOIs are the upper part of the face (from the nose to the forehead), the rest as the lower part of the face, the GUI, and the *No region*. The *No region* includes all AOIs beyond the facial image and

(a)   (b)   (c)   (d)
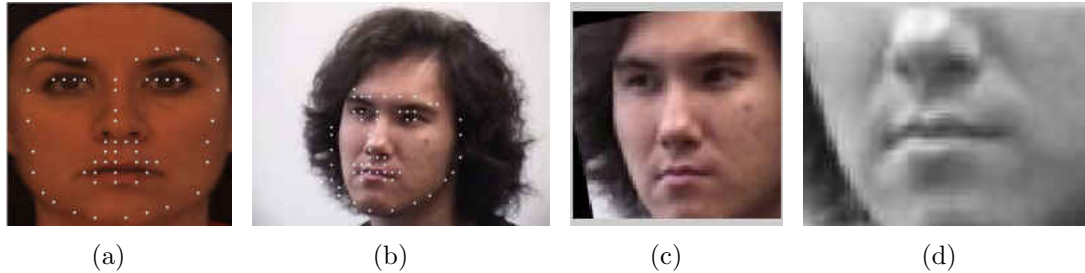
FIG. 30: (a) Ground truth reference face with facial landmarks, (b) Test face with detected landmarks, (c) Registered face after rigid transformation of landmarks, (d) Segmented lower region of the face.

the GUI. In each trial, the percentage of gaze fixation duration (GFD) spent at each of four areas is defined as follows,

$$GFD = \frac{G\_AOI}{G\_Trial} \times 100, \tag{40}$$

and is used to compare the two groups in the study. Here, G_AOI and G_TRIAL are voluntary GFDs at an AOI and the total GFD in a trial, respectively. For each subject, the percentage of GFD spent at each region is averaged over the number of trials. Within-group and between-group designs of ANOVA tests are considered with a significance level α=0.05. The ANOVA tests will reveal the significant areas of visual interest for a group during the study.

6.3 Experimental Results

Following the proposed experimental methods discussed in Section 6.3, the obtained results are presented in the following subsections:

6.3.1 Processing of Facial Image

Following the discussion in Section 6.3.3, the facial landmarks are automatically detected from color facial images as shown in Figs. 30(a) and (b). A transformation matrix is computed between two sets of landmark points corresponding to the reference face and probe face, respectively. This transformation matrix is used to perform a rigid registration of the probe face on the reference facial image as demonstrated in Fig. 30(c). Fig. 30(d) shows the segmented lower part of the face after conversion to gray scale. The lower part
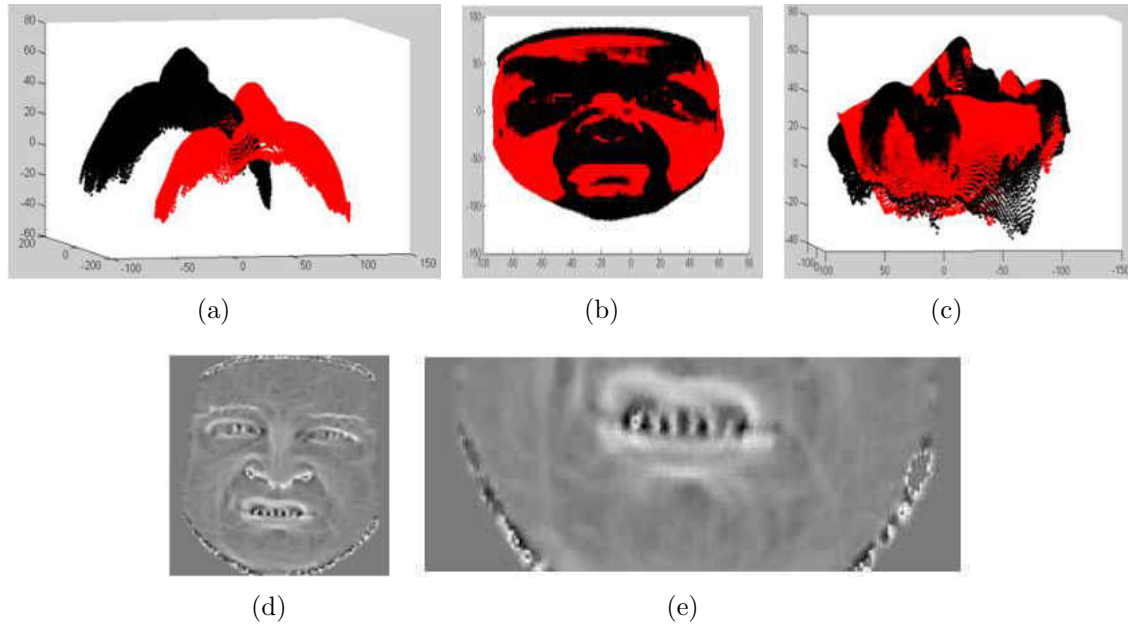
FIG. 31: ICP-based registration of 3D point cloud facial data. (a) Two 3D facial point cloud before registration, (b)-(c) After registration. (d) 2D projected surface curvature maps of 3D faces, (e) Segmented lower half of the face.

of the facial image is of 40x60 dimension, which is originally segmented from a 60x60 full facial image after the registration. The Gabor feature extraction step yields a total of 96,000 feature points from the lower part of the facial image. A small subset of representative features are selected for classification in each binary classifier following the procedure mentioned in Section 6.3.3.

Following the descriptions in Section 6.3.3, 3D facial point clouds are registered with a reference 3D facial point cloud as shown in Figs. 31(a)-(c). Following the registration, the surface-based maximum curvature is computed at each point of the interpolated 3D point cloud and projected on the 200x200 uniformly spaced grid as the intensity-based feature as shown in Fig. 31(d). The mouth region of the projected curvature feature map is segmented as shown in Fig. 31(e). The segmented lower half of the face has a dimension of 55x200, which is further downsampled to 28x100 for the ease of computation. A total of 40 Gabor filters applied over the curvature feature map yield 112,000 Gabor feature points per segmented face. This large feature dimension is reduced to a smaller set of representative features prior to classification following the feature selection step, as discussed in Section 6.3.3.

TABLE 16: 10-fold cross-validation accuracies in % for classifying five facial action units from 2D and 3D facial data.

Classifiers		Accuracy (%)	
		3D Data	2D Data
Classifier N10	AU 10 vs Neutral	95.83	95.23
Classifier N12	AU 12 vs Neutral	95	97.14
Classifier N14	AU 14 vs Neutral	95	94.28
Classifier N24	AU 24 vs Neutral	85	88.57
Classifier N25	AU 25 vs Neutral	93	96.29

TABLE 17: ANOVA test results on the percentage of prevalence of FAUs with the 2D facial data.

	Within-group design	
	Action Units	Subject
Within Control	$F(4, 35) = 10.5$ (p<0.001)	$F(7, 32) = 0.53$ (p>0.05)
Within ASD	$F(4, 35) = 6.45$ (p<0.001)	$F(7, 32) = 1.21$ (p>0.05)
Action Units	Between-group design	
AU 10	$F(1, 14) = 2.99$	p>0.05
AU 12	$F(1, 14) = 9.23$	p<0.01
AU 14	$F(1, 14) = 0.95$	p>0.05
AU 24	$F(1, 14) = 0.45$	p>0.05
AU 25	$F(1, 14) = 0.81$	p>0.05

The performance of the proposed data analysis framework is evaluated by a ten-fold cross-validation using ground truth 2D and 3D facial images from the Bosphorous dataset, respectively. Table 16 shows the ten-fold cross-validation results for the five proposed binary classifiers using benchmarked 2D and 3D facial data, respectively. The 2D facial images yield slightly better accuracies than 3D facial point cloud data except for action units 10 and 14.

6.3.2 Prevalence of FAUs in 2D Facial Data

The video camera yields more than 100 facial image frames per subject on an average. Some of the facial image frames are discarded in which the faces are partially out of the

FIG. 32: Mean prevalence of FAUs and standard error plots for 2D facial images. (a) Subject-wise mean prevalence of all FAUs for subjects with ASD, (b) Subject-wise mean prevalence of all FAUs for TD subjects, (c) Mean prevalence of different FAUs in two groups, (d) Overall mean prevalence of FAUs in two groups of subjects.

frame, ill-posed, or occluded because of the unconstrained head motion. The trained classifier models are used to detect the prevalence of different FAUs from the test facial images obtained from the study. Figure 32(a) shows that the average percentage of prevalence (APP) of FAUs, measured over all FAUs, is below 40% for the TD subjects. However, the subjects with ASD demonstrate variability in the APP measured over all FAUs, which is over 40% for several subjects as shown in Fig. 32(b). A within-group design ANOVA reveals that the TD subjects have no significant difference among themselves, $F(7, 32)=0.53$, ($p>0.05$) as shown in Table 17. Similarly, the subjects with ASD reveal no significant difference in terms of the APP measured over all FAUs, $F(7, 32)=1.21$, ($p>0.05$). Within

a group, further ANOVA tests show significant difference among APPs of the five FAUs. The APPs of five FAUs have been found to be significantly different for the TD group, $F(4, 35) = 10.5$, $(p<0.001)$ as well as for the group with ASD, $F(4, 35) = 6.45$, $(p<0.001)$. Between-group ANOVA tests are performed to investigate the difference between the two groups. Table 17 summarizes the between-group ANOVA test results, which show no significant difference between two groups in terms of APP of FAUs except for FAU 12. Figure 32(c) provides a comparative illustration of the two groups in terms of APPs related to five FAUs, respectively. The ASD group exhibits a significantly higher prevalence of FAU 12, $F(1, 14) = 9.23$, $(p<0.01)$ than the TD group, which is evident in Fig. 32(c). A comparative illustration between the TD and ASD groups is shown in Fig. 32(d) in terms of the APPs of all FAUs measured over all subjects within a group. This overall illustration reveals that the group with ASD (37.21%) manifests a higher APP of FAUs when compared to the TD group (27.94%).

6.3.3 Prevalence of FAUs in 3D Facial Data

The 3D optical camera reconstructs 3D facial point cloud data from multiple stereoscopic images, which are sensitive to the positioning of the subject's head. If a stereo image misses a portion of a facial side, it may distort the final reconstructed 3D face. There have been subjects with ASD who are restless in nature, and therefore sometimes fail to stay within the camera's field of view. In addition to the sensor limitation as mentioned earlier, a number of 3D facial images have been discarded for incomplete or distorted reconstruction. Therefore, on an average, about 12 3D-facial images are obtained per subject. The following section discusses the prevalence of FAUs for 3D facial images collected in this study.

Table 18 summarizes the results following the ANOVA tests performed using 3D facial point cloud data. A within-group design ANOVA test shows no significant difference among the subjects within the TD group in terms of their APP measured over all FAUs, $F(7, 32) = 1.02$, $p > 0.05$ as illustrated in Fig. 33(a). However, the subjects with ASD have been found significantly heterogeneous in terms of APP measured over all FAUs, $F(7, 32) = 14.46$, $p < 0.001$ as shown in Fig. 33(b).

Table 18 shows that there is no significant differences among five FAUs in terms of their APPs measured over all subjects within the group with ASD, $F(4, 35) = 0.46$, $p > 0.05$. However, the TD group reveals a significant difference among their elicited five FAUs in terms of APP, $F(4, 35) = 4.21$, $p < 0.01$. A between-group design ANOVA test is
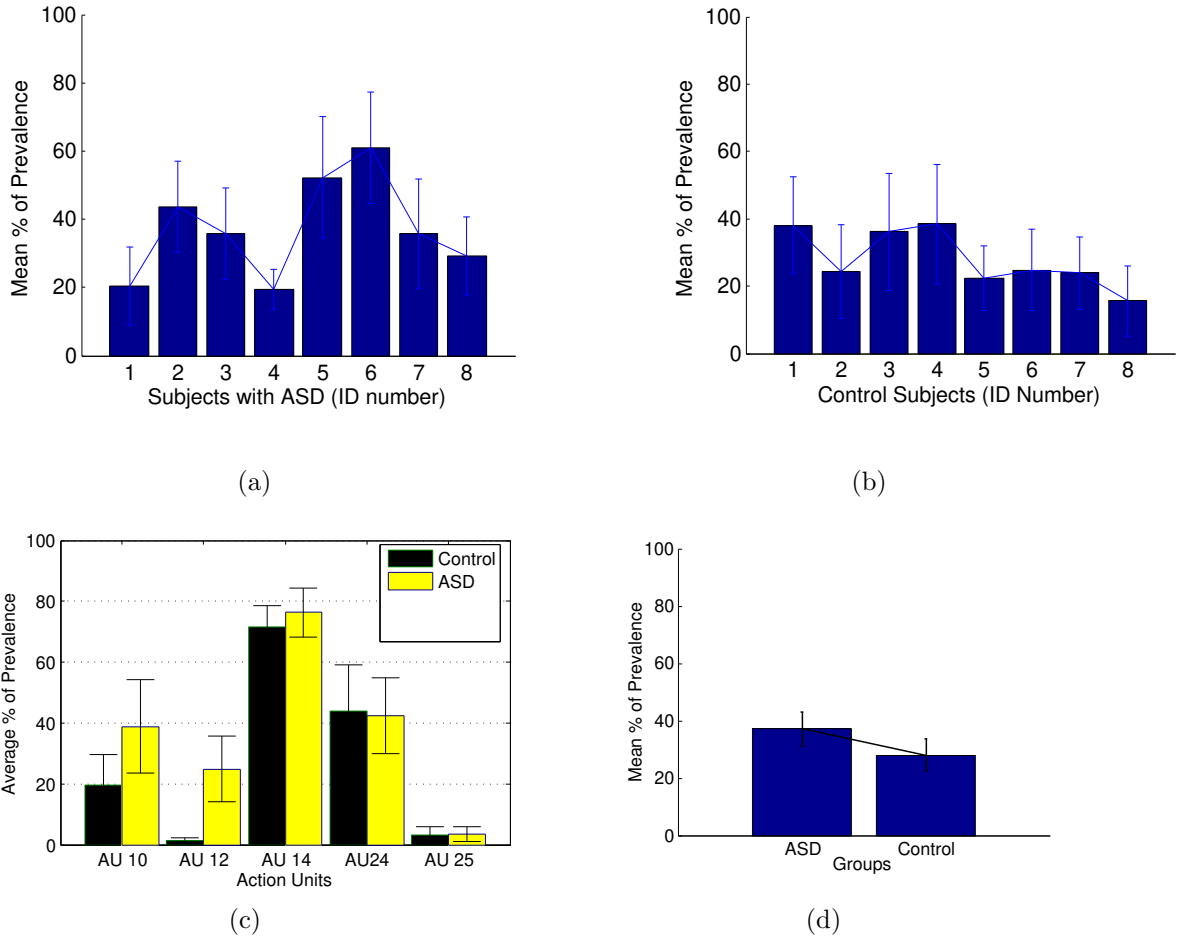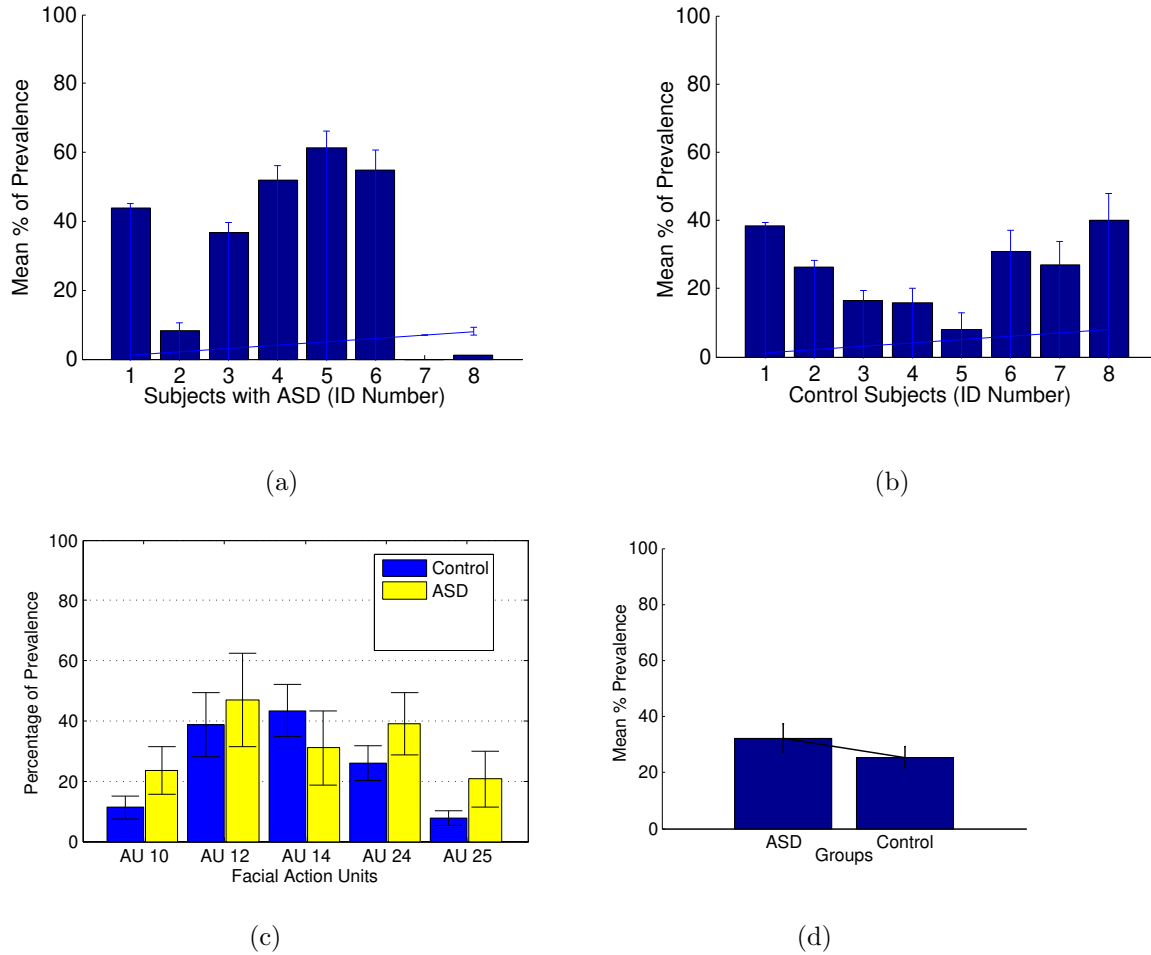
FIG. 33: Mean prevalence of FAUs and standard error plots for 3D facial images. (a) Subject-wise mean prevalence of all FAUs from subjects with ASD, (b) Subject-wise mean prevalence of all FAUs from control subjects, (c) mean prevalence of different FAUs in two groups (d) Overall mean prevalence of FAUs in two groups of subjects.

performed to identify the statistical difference between the two group in terms of the APP measured for each FAU. Table 18 shows that there is no significant difference between ASD and control groups in terms of APP of any FAU, which is also illustrated in Fig. 33(c). Similar to the observation with 2D facial images, the APP measured over all subjects and all FAUs for the group with ASD (32.27%) is found higher than that for the TD group (25.37%) as shown in Fig. 33(d). Note that the 3D facial point cloud data do not yield a statistically significant result in between-group ANOVA tests as in the case of 2D facial images. Therefore, we further discuss the subject-specific findings about the percentage of prevalence of different FAUs.

TABLE 18: ANOVA test results on the percentage of prevalence of FAUs for 3D facial data.

	Within-group design	
	Action Units	Subject
Within Control	$F(4, 35) = 4.21$ ($p<0.01$)	$F(7, 32)=1.02$ ($p>0.05$)
Within ASD	$F(4, 35) = 0.46$ ($p>0.05$)	$F(7, 32) = 14.46$ ($p<0.001$)
Action Units	Between-group design	
AU 10	$F(1, 14)=0.3$	$p>0.05$
AU 12	$F(1, 14)=0.78$	$p>0.05$
AU 14	$F(1, 14)=0.12$	$p>0.05$
AU 24	$F(1, 14)=0.82$	$p>0.05$
AU 25	$F(1, 14)=0.91$	$p>0.05$

Fig. 34 demonstrates the subject-specific percentage of prevalence for different FAUs. Fig. 34(a) shows that the subjects with ASD labeled by ID numbers 1, 3, 4, and 6 have more prevalence of AU 10 than their TD peers. Fig. 34(e) shows a high prevalence of AU 25 with the same subjects as for AU 10, since 'upper lip raiser' (FAU 10) usually co-occurs with 'mouth open' (AU 25) involving the activation of *Orbicularis Oris* muscle around the lip region.

Subjects with ASD, labeled 4, 5, and 6, have noticeably produced a high prevalence of AU 12, which represents 'lip stretching' as the main component of smile expression as shown in Fig. 34 (b). Fig. 34 (c) shows high prevalence of 'facial dimple formation' (AU 14) with the same subjects. This is because facial AU 14 is activated by *Buccinator* muscle, which can co-occur with the activation of muscle *Zygomaticus Major* resulting in AU 12. Subjects with ASD, labeled 2, 7, and 8, reveal a high prevalence of neutral expression or lack of facial affect, except a few prevalence of 'lip pressing' represented by AU 24 (Fig. 34 (d)). On the other hand, TD subjects with labels 1 and 8 have a moderate prevalence of all FAUs. TD subjects with labels 5, 6, and 7 have a moderate presence of all FAUs, except AU 10 and AU 25, which may appear together as mentioned before. TD subject 4 excludes AU 25, subject 2 excludes AU 10, and subject 3 excludes both AU 10 and AU 12. Fig. 34 (f) shows the APP over all five FAUs for individual subjects belonging to the two groups. This reveals that subjects with ASD have been found facially more active than their TD peers.
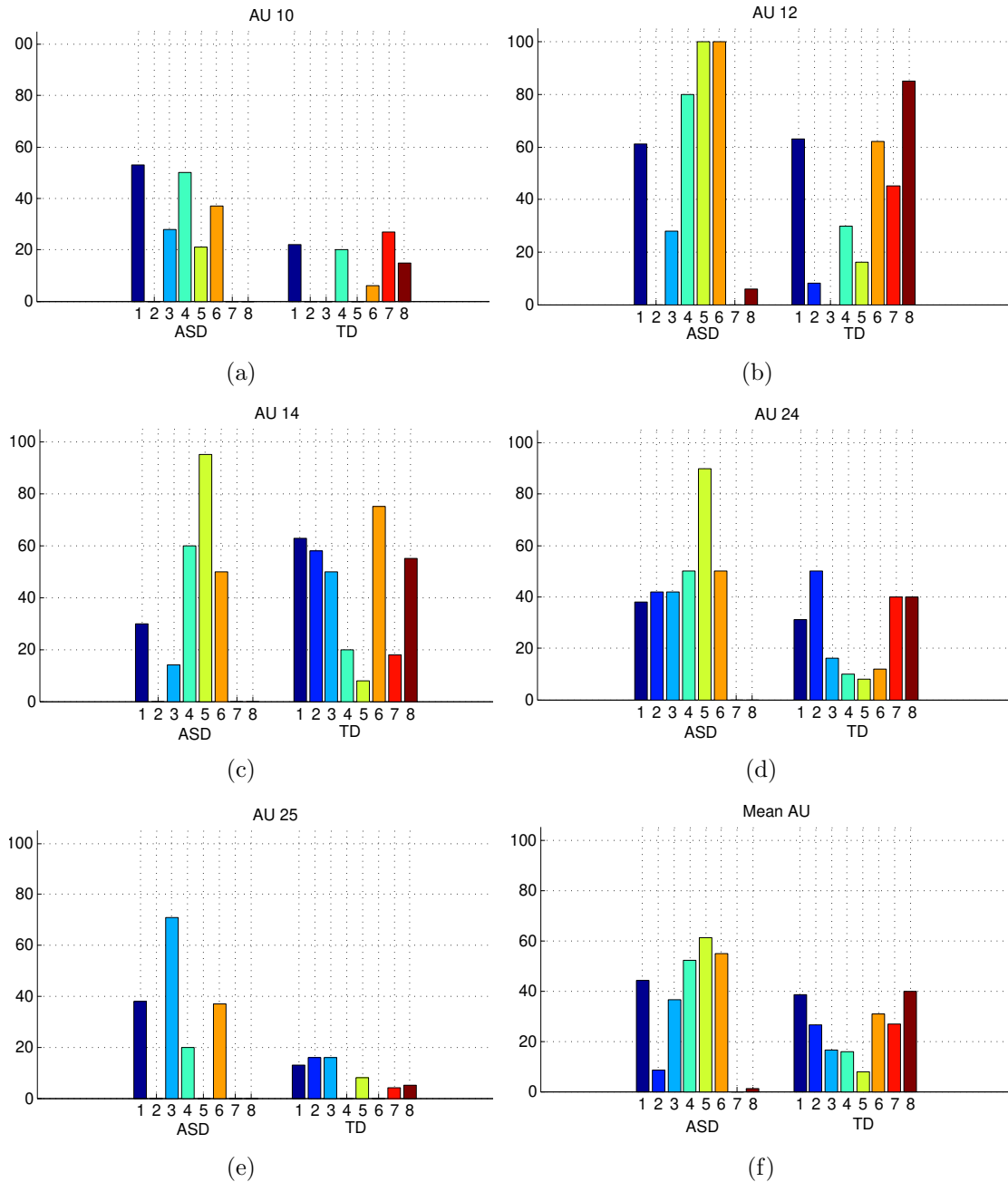
FIG. 34: Subject-specific prevalence of five different FAUs in eight subjects with ASD and eight TD subjects. Percentage of prevalence of (a) AU 10, (b) AU 12, (c) AU 14, (d) AU 24, (e) AU 25, (d) Mean FAUs.

6.3.4 Results from the Eye-tracker data

Following the procedures in Section 6.3.4, the percentage of gaze fixation duration (GFD) spent at each visual AOI is computed for individual subjects. The percentage of

TABLE 19: ANOVA test results. The upper triangle is for the group with ASD, the lower triangle is for the TD group. The significant inferences are highlighted. No = *No region*, UP = upper part, LP = lower part of the face, GUI = graphical user interface, PVGD = percentage of voluntary gaze duration.

Within-group design			
	No	UP	LP	GUI
No	X	F=19.95, **p<0.001**	F=13.72 **p<0.001**	F=1.4, p>0.05
UP	F=2.12 p>0.05	X	F=0.54 p>0.05	F=74.21 **p<0.001**
LP	F=3.63 p>0.05	F=6.64 **p<0.05**	X	F=48.21, **p<0.001**
GUI	F=23.26 **p<0.001**	F=1.48 p>0.05	F=43.22, **p<0.001**	X
Between-group design			
PVGD	No	UP	LP	GUI
F=4.06, p=0.063	F=6.39, **p<0.05**	F=7.54, **p<0.05**	F=0.35, p>0.05	F=1.3, p>0.05

time spent on voluntary GFD over all GFD (both voluntary and involuntary) has been found to be higher within the TD group (74%) when compared with the group with ASD (68%), however, this difference is not statistically significant (F(1, 14)= 4.06, p>0.05). This is shown as the percentage voluntary gaze duration (PVGD) in Fig. 35 and Table 19. The following sections describe the findings from the ANOVA tests.

Fig. 36 shows an example of overall gaze maps in a trial for two age-matched subjects (one with ASD and one TD subject). It is obvious from the gaze map that unlike the TD subject, the subjects with ASD avoided the upper part of the face and rather gazed at the lower part and elsewhere.

The ANOVA test results reveal some interesting observations in Table 19. The upper triangle of Table 19 reveals that the group with ASD has spent a significantly higher percentage of GFD in scanning the GUI and the *No region* than at the facial regions (lower and upper face). A significantly higher percentage of GFD is found at the *No region* when compared with the upper part (F=19.95, p<0.001) and the lower part (F=13.72, p<0.001) of the face. Similarly, the ASD group has spent significantly more percentage of GFD in scanning the GUI than the upper (F=74.21, p<0.001) and lower (F=48.21, p<0.001) parts of the face. However, no significant difference is found between the percentages of GFD
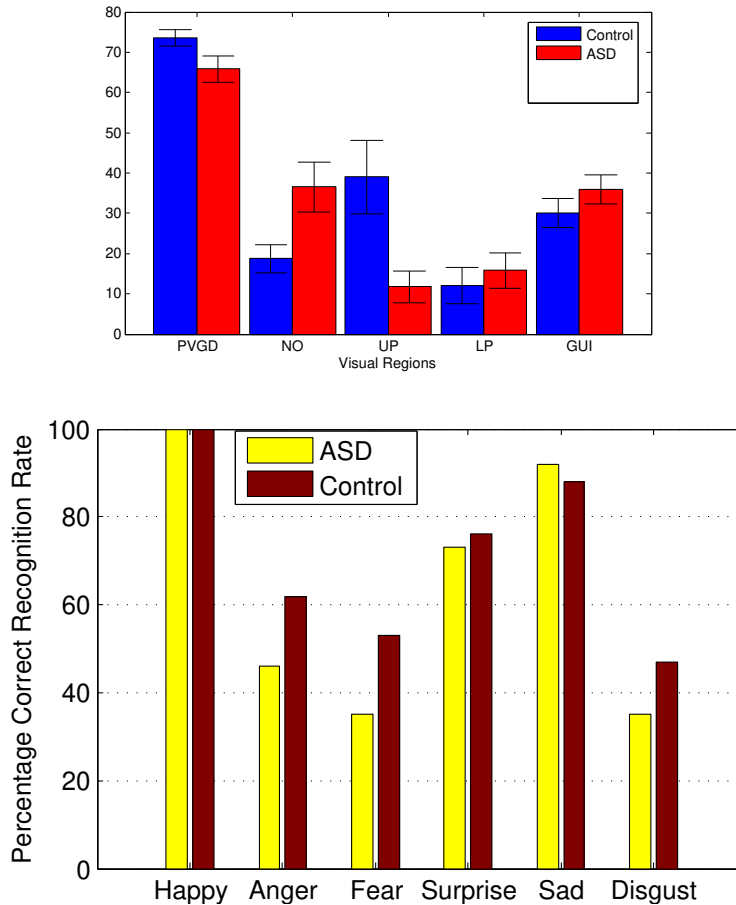
FIG. 35: (a) Percentage of gaze duration at four visual scenes (No = *No* region, UP = upper face, LP = lower face, GUI = Graphical User Interface). PVGD = Percentage of Voluntary Gaze Duration.

spent in scanning lower and upper parts of the face (F=0.54, p>0.05) respectively. This indicates that individuals with ASD have a tendency to avoid visual interaction with the human face and exert more visual scanning on elsewhere. The findings are quite different for the TD group as shown in the lower triangle of Table 19. The TD group has spent a significantly higher percentage of GFD in scanning the upper part of the face than the lower part (F=6.64, p<0.05). Individuals in the TD group have spent significantly less percentage of time in scanning the *No region*, (F=23.26, p<0.001) and lower part of the face (F=43.22, p<0.001) than that spent in scanning the GUI. A between-group ANOVA test indicates an insignificant difference between two groups in terms of the percentage of time spent in scanning the GUI (F= 1.3, p>0.05) and lower part of the face (F=0.43, p>0.05) as shown in Table 19. However, the TD group has spent a significantly higher

(a)   (b)
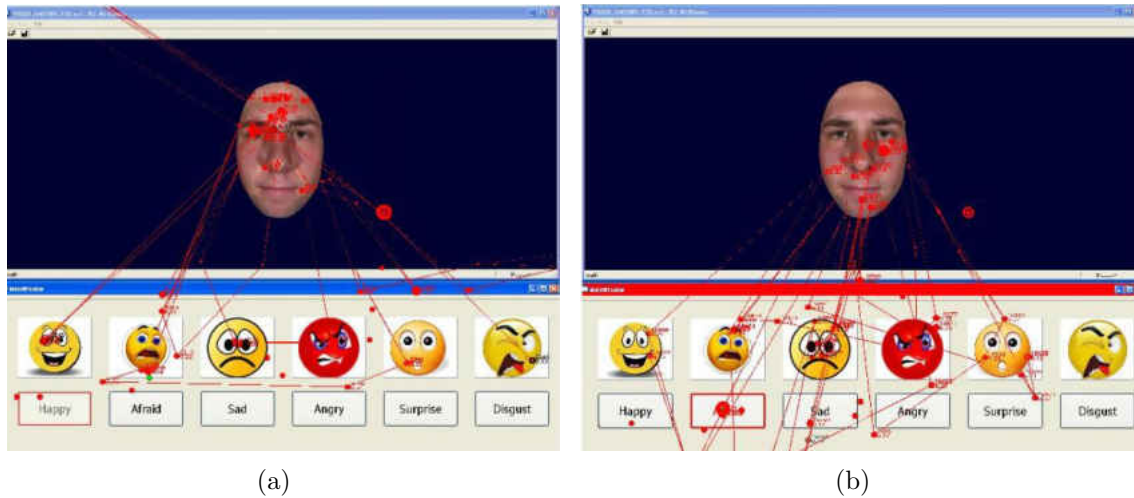
FIG. 36: Gaze pattern in an attempt to recognize a facial expression. Gaze patterns produced by (a) A TD subject and (b) A subject with ASD.

percentage of time visually scanning the upper part of the face than that spent by the group with ASD, (F=7.54, p<0.05). The group with ASD has spent a significantly higher percentage of time scanning the *No region* than that spent by the TD group, (F=6.39, p<0.05).

6.3.5 Perception of Emotions

The facial expression recognition task in Session 1 of the first experiment also examines the subjects perception of the displayed emotional expression, which is collected on the GUI. Fig. 35(b) shows the percentage correct recognition of six facial expressions for two groups of subjects. Similar to a previous study [122], Fig. 35(b) reveals that subjects with ASD have performed poorly in recognizing negative expressions such as anger, fear, and disgust. This difference indicates that subjects with ASD may fail to grasp a social context due to the impairment in their perception of negative facial expressions. This lack of perception of emotions may, in turn, regulate or limit their facial responses as found in the previous sections.

6.4 Discussion of the Results

The quantitative evaluations in this chapter demonstrate several differential responses that can provide important insights into the symptoms and severity of the group with

ASD. The analysis of FAUs from videotaped data reveals a significant prevalence of FAU 12, which is the major constituent of smile expression. Although the smile expression is, in general, visible, any anomaly pertaining to the smile expression (FAU 12) at an early developmental age can be subtle for human eyes. Smiling and crying are known to be the earliest facial responses of human infants as classified by baby FACS [123]. The anomaly in the prevalence of smile expression for the school-aged group with ASD may also be a target for screening younger children with ASD at preschool ages. An early screening and identification of ASD-related traits will, in turn, enable early and effective intervention planning.

Note that the higher prevalence of FAUs in the group with ASD when compared with the TD group (See Fig. 7(d)) may contradict a general notion that subjects with ASD are unable to produce facial expressions. However, this study enrolls high-functioning subjects with ASD who are known to engage and respond more appropriately than their low-functioning peers when the stimulus or the task is arousing [53] [124] [16]. The results in Chapter 4 involving muscle-specific analyses of both 2D and 3D facial data also reveal higher facial muscle activation in the group with ASD when compared with the TD group [23]. From the definition of FACS, the activation of *Zygomaticus Major* muscle may be studied using the APP of FAU 12 [89]. Unlike prior studies [54, 60, 46] with intrusive methods and undifferentiated findings, the proposed non-intrusive computer vision-based method shows a significantly higher prevalence of FAU 12 for the group with ASD.

The FAU 12 is the principal component of smile expression, whereas social smile and facial expressions are clinically screened in individuals with ASD [14]. Social smile is a reciprocal behavior that involves eye contact during sharing of smile between two individuals. In Section 6.4.4, the eye-tracker data reveal that the subjects with ASD have spent significantly more time elsewhere than visualizing the faces with expressions. The high prevalence of smile expression along with such averted gaze patterns violates the reciprocity of social smile. Therefore, the impairment in social smile in subjects with ASD may be automatically computed simultaneously from facial responses and eye-gaze patterns using non-intrusive computer vision sensors. Furthermore, human infants can manifest social smile as early as six months of age by making face-to-face interaction with adults [125]. If the impairment in social smile is identified earlier, it may consequently provide an early target for behavioral intervention. The efficacy of the behavioral intervention to mitigate the anomaly in social smile may further be evaluated using the proposed computational methods.

The averted eye-gaze patterns may have contributed to poor emotion recognition performance of the subjects with ASD as shown in Section 6.4.4. The lack of perception of negative emotional expressions such as anger, fear, and disgust may have favored the manifestation of smile incongruent to the visual stimuli for these subjects. Smiles incongruent to other channels of verbal or nonverbal communication such as vocal prosody, eye-gaze and gesture may be an effective behavioral marker for ASD. The subjective evaluations of human raters in prior studies report atypical traits in the facial responses of subjects with ASD when compared with the TD group [126, 16, 17]. An objective evaluation via computer vision-based recognition of FAUs and eye-tracking data, as proposed in this study, may offer necessary metric to quantify the atypical behavioral traits from the elicited facial and visual responses.

In contrast to the group with ASD, the low prevalence of FAU 12 in TD subjects may be due to a lack of enthusiasm in the tasks with facial expressions presented via a computer-simulated environment. The eye-tracker reveals that TD subjects are visually focused on the upper part of the face (eye-contact region) and GUI while attempting to provide the answer. This suggests that the TD group may have more control of their emotional expressions than the ASD group. Individuals suffering from neurological disorders like ASD are known to have less control over their emotional expressions [127]. Emotional expressions such as smiling and crying can be induced involuntarily even in cases of palsy or lack of emotional control. This is because facial nerves originating from the thalamus regulate involuntary control of facial muscles, which may elicit involuntary smile expression [48]. Furthermore, prior studies report that subjects with ASD are more comfortable and engaged during a computer-based interface or virtual environments than during real life human interactions [104, 105]. Therefore, similar computer-based virtual environments may be used to stimulate and intervene the anomalies detected as behavioral markers.

This chapter proposes a computational model to investigate the behavioral markers that differentiate the group with ASD from the TD group. The behavioral markers are quantitatively studied from the elicited facial responses, eye-gaze patterns, and perceptual skills of the subjects. These spontaneous responses are common targets for follow-up visual screenings of children with ASD. The FACS-based encoding of facial responses and the analysis of gaze patterns reveal that the group with ASD manifests intense and frequent smile expression with an averted gaze pattern. The lack of congruence in facial responses and eye-gaze suggests a behavioral marker, which may be useful in quantifying nonverbal skills of subjects with ASD.

CHAPTER 7

DYNAMIC FACIAL RESPONSE IN ASD

7.1 Chapter Overview

This chapter discusses the second experiment as outlined in Section 3. The second experiment employs a motion capture sensor to acquire dynamic facial actions in response to dynamic audio-visual stimuli as discussed in Section 3.3.1. The narrative of the story is provided in Appendix C. The time-sampled dynamic facial response data are encoded using FACS-based FAUs. The differential traits for the group with ASD are investigated from the prevalence of different FAUs in the time-sampled facial action data in response to the stimuli. Several metrics are proposed and investigated to compute the differential traits for the group with ASD using the dynamic facial action data. Within-group and between-group ANOVA tests are performed to statistically evaluate the prevalence of differential traits in the facial expressions of the group with ASD compared to the TD group using dynamic facial response data.

7.2 Analysis of Dynamic Facial Response

For the purpose of this study, ten FAUs are considered corresponding to ten distinct facial actions detected and tracked by *faceshift* as shown in Fig. 37. These action units cover a range of facial actions from the eye-brow to the jaw. Du et al. delineate the constituents of different compound emotional states in term of FAUs [91]. The proposed ten FAUs contributing to different compound emotional expressions are shown in Table 20. The prevalence of the ten target FAUs were computed and compared for the two groups to identify several differential traits in the spontaneous facial responses. The differential traits associated with the group with ASD are computed using the following three metrics.

First, the number of times a FAU is activated is counted for each subject in response to the four-minute long audio-visual stimuli. The numbers of activation within a group are averaged over all ten subjects for each FAU. This yields the mean and standard error of the prevalence of a particular FAU within a group. The mean prevalence of a FAU is utilized to compare between the TD and ASD groups using a t-test. This analysis reveals

TABLE 20: FACS-based annotation of facial actions tracked by the facial motion capture system (*faceshift*).

Facial Actions (*faceshift*)	Facial Action Units (FAUs)	Representative Emotional Expressions
Brows Up	AU 1	Sadness, Surprise Fear
Sneer	AU 6	Happiness, Scorn
Upper Lips Up	AU 10	Disgust
Smile	AU 12	Happiness
Dimple	AU 14	Depression [93]
Mouth Frown	AU 15	Disgust, Sadness
Lower lip down	AU 16	Disgust
Jaw	AU 17	Disgust
Lips stretch	AU 20	Fear
Mouth Press	AU 24	Anger

any deficit in the mean prevalence of a FAU in the group with ASD when compared with that for the TD group.

Second, the effects of emotional context of the audio-visual stimuli on the prevalence of FAU are computed. A story narrated by an Avatar is used as the audio-visual stimuli (See Appendix C). The story demonstrates multiple emotional states of the Avatar and shows contexts as categorized in Table 21. For each of the ten stimuli contexts (labeled with stimuli ID), the average number of activations for all FAUs per subject is computed. Within a group, the numbers of activation are averaged over all ten subjects for each stimuli context (stimuli ID). This yields the mean and standard error of the prevalence of the FAUs within a group for a particular stimuli context.

Third, the correlation coefficient is computed between the time-sampled activation data of two FAUs in response to the four-minute long audio-visual stimulus. A prototypical facial expression (e.g. happy, anger, or surprise) may be a combination of multiple FAUs. Therefore, the facial muscle actions are likely to be correlated, where the activation of one FAU can be tied to the activation of another. Between-group ANOVA tests are conducted to ascertain if any group has significantly higher correlation coefficient for a pair of FAUs than that of the other group. The lack of correlation between a pair of FAUs suggest a measurable differential trait in the facial physiology for the group with ASD.
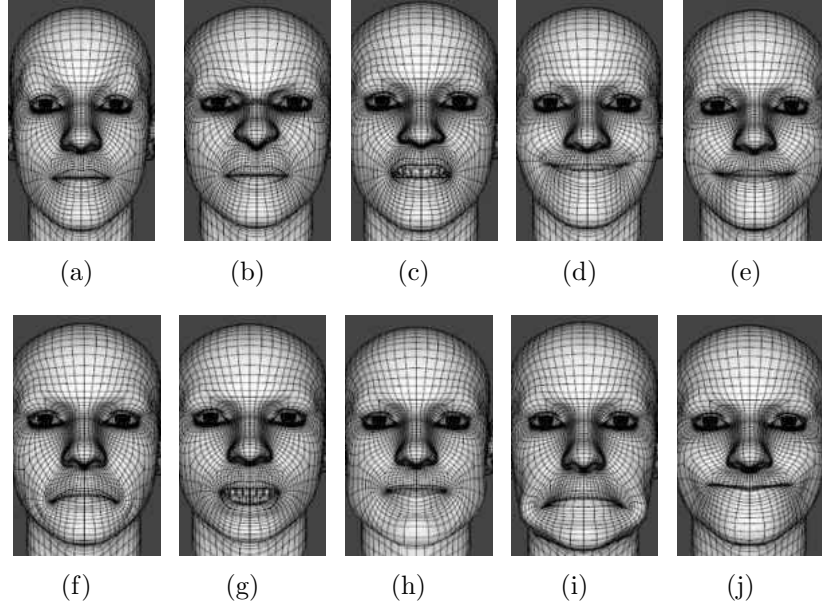
FIG. 37: Synthetically generated ten facial action units on 3D deformable facial model for visualization. (a) FAU 1, (b) FAU 6, (c) FAU 10, (d) FAU 12, (e) FAU 14, (f) FAU 15, (g) FAU 16, (h) FAU 17, (i) FAU 20, (j) FAU 24

7.3 Results

The section describes the results obtained following the methods proposed in Section 7.2.

7.3.1 Mean Prevalence of Facial Action Unit

The magnitude of activation for a FAU is normalized between the values of 0 and 1. The threshold for activation is set equal to or higher than 0.3 by trial-and-error to represent the activation of the FAU at a time instance. The magnitude of activation data for each FAU are collected via the *faceshift* application at a rate of 24 samples per second. The four-min time series data of facial activation in response to the four-min audio-visual stimuli are downsampled to 240 sample points per FAU. We count the number of the activation of each FAU and average over ten subjects within a group. Fig. 38 shows the mean activation of each FAU over ten subjects for each group. Facial action units 1, 14, 16, 17, and 24 appear to yield low activation count when compared with the other five action units. The TD group appears to yield higher activation of FAU 6 (sneer), FAU 12 (smile), and FAU 20 (lip stretch) when compared with the group with ASD. Interestingly, the group with

TABLE 21: Temporal breakdown of the proposed four-minute long Avatar-based audio-visual stimuli (narrated in Appendix C) for context and emotional content.

Stimuli ID	Context of Stimuli	Total Time(in sec.)	Avatar Emotion
1	Introduction: Introduce School and Friends	40	Neutral
2	Plan for watching movie with friends	14	Excited
3	Pizza and anchovy in Pizza	16	Disgusted
4	Tony took my notebook and forgot to return	29	Angry
5	Missing notebook, test tomorrow!	28	Anxious & Sad
6	Tony shows up with notebook and a candy bar	13	Surprise
7	In the Exam Hall with the test and a candy bar	25	Tensed
8	Saturday Bowling party	12	Happy
9	Bowling Performance	42	Suspense
10	Saying good bye	7	Neutral

ASD has produced a higher mean activation of FAU 15 (mouth frown) when compared with the TD group. However, none of these differences is statistically significant.

Since five FAUs (1, 14, 16, 17, 24) have low yields in terms of activation, these action units are excluded from the within-group design ANOVA tests. A post-hoc Tucky test is performed on the activation of FAU pairs within a group. The five FAUs (6, 10, 12, 15, 20) that are the most active for both groups are considered. For the group with ASD, there is no significant difference among the activations of five FAUs, $F(4, 45) = 0.54$, $p > 0.05$. Further post-hoc Tucky test reveals no significant difference between the activations of any two FAUs. For the TD group, there is no significant difference among the activations of FAUs, $F(4, 45) = 1.21$, $p > 0.05$. However, a post-hoc Tucky test shows a significantly lower activation of FAU 15 (mouth frown) when compared with that of FAU 12 (smile), [F(1,
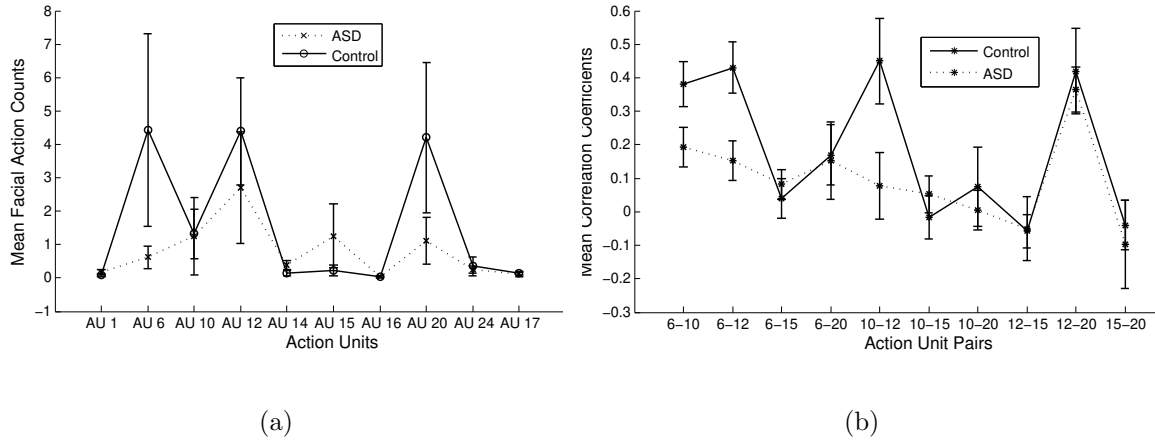
FIG. 38: (a) Mean count of ten different facial action units from time-sampled data averaged from ten subjects in each group. (b) Mean correlation coefficients over ten subjects for each pair of facial action units. The bar indicates standard error.

18)=6.74, p<0.05] for the TD group. This is also evident in Fig. 38(a), which shows an absence of mouth frown (AU 15) in the TD group unlike the group with ASD.

7.3.2 Correlation Coefficients of facial action units

The correlation coefficient for each pair of time-sampled FAU activation data is computed for each subject. The five most prevalent FAUs (6, 10, 12, 15, 20) are considered in this case. The computed correlation coefficients are further evaluated in between-group ANOVA tests. The five FAUs make a total of ten pairs of facial actions. Fig. 38(b) shows the correlation coefficients for ten pairs of FAUs units averaged for ten subjects in each group. Between-group ANOVA tests reveal that the TD group has significantly higher correlation coefficients for three FAU pairs than those with the group with ASD. These FAU pairs and corresponding statistics are: FAU 6 and AU 10, [$F(1,18)=4.24$, $p=0.05$], FAU 6 and FAU 12, [$F(1,18)=7.98$, $p<0.05$], FAU 10 and AU 12 [$F(1,18)=5.08$, $p<0.05$]. The correlation coefficients for the other pairs of FAUs yield insignificant difference between the two groups. These results indicate that unlike the TD group, the group with ASD has significantly lower correlation in the activation of FAUs. The lack of correlation in the activation of facial actions may be a potential anomaly that suggests a differential trait for the subjects with ASD.
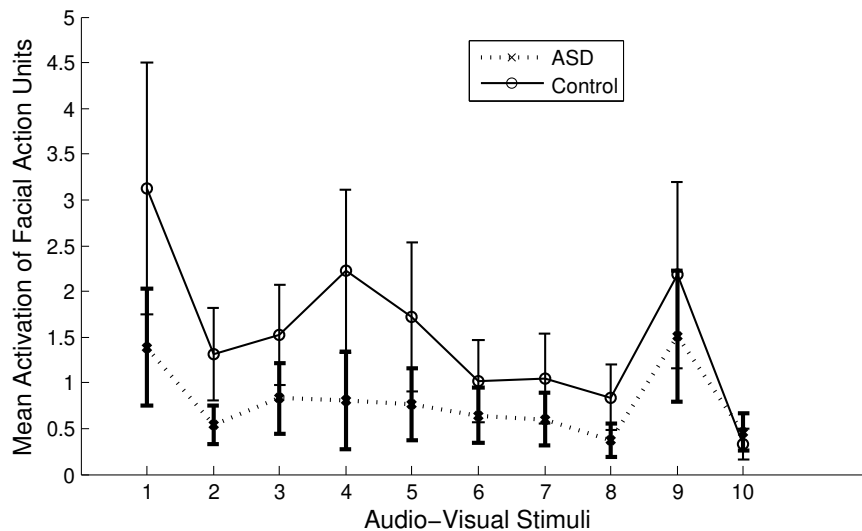
FIG. 39: Mean count of FAUs for the group with ASD and the TD group. Effects of different types of stimuli context on facial responses averaged from all facial action units.

7.3.3 Stimulus-specific facial response

Fig. 39 shows mean activation plots of all FAUs for the group with ASD and the TD group. The mean activations of all FAUs are higher for the TD group than those for the group with ASD. The changes in the mean activations over the stimuli ID appear to be similar for both groups except at the periods of stimuli IDs 4 and 5. Table 21 shows that these two stimuli refer to the contexts when the Avatar demonstrates angry, anxious, and sad expressions. At this point of the story, Tony forgets to return Robert's (Avatar character) notebook, the exam is scheduled tomorrow, and Robert has no other means to study or prepare. The flatness in the mean activation patterns of FAUs for the group with ASD indicates that unlike the TD group, the individuals with ASD tend to be apathetic to the state of anger and anxiety of their peers.

7.4 Discussion of Results

In this chapter, the differential traits in the spontaneous facial expression of the group with ASD are objectively computed using three metrics: unusual prevalence of an odd FAU (e.g., mouth frown, FAU 15), the lack of correlation in the time-sampled responses of several FAU pairs, and lack of response to emotional trigger (stimuli). Although the FAU

15 (mouth frown) is almost absent in the TD group, it is detected at a higher magnitude in the group with ASD. Unlike the TD group, the presence of mouth frown (FAU 15) may have diminished the natural traits in the expression, as noticed by human raters employed in the prior studies [16, 17]. On the other hand, the smile expression constitutes both FAU 6 and 12. A variety of emotional states can be represented by the smile expression apart from happiness and the time-sampled data of smile expression can be a useful target in the investigation of covert emotional states [47]. Whereas both FAUs of the smile expression are equally prevalent in the TD group, the group with ASD shows a flatness in the activation for the FAU 6 than that of the FAU 12 (See Fig. 38(a)). Such flatness in the prevalence of FAU 6 may be a differential trait in the smile expressions of the group with ASD. Furthermore, the smile expression should yield the activation of FAU 6 and FAU 12 simultaneously with a higher correlation in their time-sampled activation data. However, the group with ASD yields significantly lower correlation coefficients between the time-sampled activation data for FAU 6 and FAU 12 unlike the TD group. The significant absence of correlation in the time-sampled activation data for different FAUs may be a measurable differential trait for the individuals with ASD. All these results indicate obvious deviations from the typical functioning and prevalence of the relevant FAUs typically found with the TD group.

The group with ASD is also known to have deficits in the perception about the emotional and social contexts [122, 44]. In behavioral studies, such deficits are investigated subjectively via visual screening [56]. This deficit may be identified quantitatively using the stimuli-specific activation data of FAUs as shown in this study. Therefore, the lack of facial response at the event of emotional trigger may be a measurable trait for subjects with ASD. Fig. 39 shows that on an average the group with ASD yields less activations of FAUs for all contexts and emotional trigger displayed by the audio-visual stimuli. This reveals a lack of empathy of the subjects with ASD toward emotional contents and contexts of the stimuli. Therefore, the proposed study procedure and computational models may be used for quantitatively screening of subject-specific empathy in response to an emotional trigger.

In summary, this chapter demonstrates a novel method to compute the differential traits from the spontaneous and dynamic facial motion capture data. The FACS-based annotation of the time-sampled facial activation data offers a fine-grain quantitative analysis of spontaneous facial response in the measurement of differential traits for the subjects with ASD. The dynamic facial actions in response to dynamic audio-visual stimuli are more realistic than using the static stimuli as discussed in Chapter 3 for the first experiment.

CHAPTER 8

SUMMARY AND FUTURE WORK

This dissertation proposes experiments and computational models for quantitative psychophysical study of spontaneous facial expressions of subjects with ASD in a non-intrusive manner. The work herein demonstrates several novel computational models and computer vision-based frameworks in the investigation of differential traits from the spontaneous facial response of subjects with ASD. Following IRB-approval, two psychophysical studies have been conducted to collect and analyze multimodal data from for subjects with ASD and a control group with TD subjects. The spontaneous facial expressions capture the actual psychophysical response unlike the study of posed or imitated facial expressions in the literature. The proposed methods alleviate the existing drawbacks in the literature about subjective and intrusive procedures to study the differential traits in the facial expressions of subjects with ASD. Consequently, this dissertation proposes a number of quantitative and objective methods and metrics for extensive analysis of psychophysical response from facial imaging data.

Extensive statistical analyses and comparison of psychophysical response data from both groups reveal the differential traits in the facial expressions for the group with ASD. These differential traits may suggest several biomarkers for the group with ASD. The overall contributions of this dissertation are summarized in Table 22 and are further discussed below.

First, a curve-based representation of 3D facial data is proposed and an efficient computational model is developed to extract pose-invariant geometric features from the curve-based representation in Chapter 4. The proposed computational models and geometric features offer several benefits such as pose robustness, competitive performance in face and facial expression recognition, reduced cost of computation, and facial muscle-specific localization, and analysis of 3D facial data. Second, the proposed curve-based representation of the 3D face has shown promising results in the non-intrusive computation of asymmetry in the facial muscle actions. Chapter 5 shows statistically significant asymmetry in the facial muscle actions for the group with ASD in contrast to the TD group. This offers a non-intrusive computational model to objectively compute the differential traits in

TABLE 22: Summary of the research findings related to the proposed methods and the group with ASD.

Chapter	Proposed Methods	Description	Outcomes
4	Computational modeling of Frenet frame-based novel features	Competitive performance for 3D face & facial expression recognition under pose variations	Pose-invariant & faster 3D facial analysis & recognition [21, 22]
5	3D facial data for computing facial muscle activation	Significant asymmetry in left-right facial muscle actions for subjects with ASD	A metric to compute oddity in the facial expressions [23, 24]
6	FACS-based facial response analysis from 2D/3D facial data	Significant prevalence of FAU 12 for subjects with ASD	Uncontrolled facial expression for subjects with ASD [25]
6	Eye-tracking while visualizing 3D faces with expressions	Significantly averted gaze & eye-contact for subjects with ASD	Poor reciprocity in social smile Poor perception of negative expressions
7	Facial MoCap sensor to study dynamic activation of FAUs in response to social stimuli	Prevalence of mouth frown, lack of correlation in FAU actions, flatness in FAU with subjects with ASD	Computing of traits and apathy using dynamic facial response data [26]

facial expressions of individuals with ASD, which is reported in qualitative studies on ASD . Third, the dissertation proposes FACS-based classification of subtle spontaneous facial expressions for the first time in ASD literature. The computer vision-based classification of FAUs using video data reveals significantly higher prevalence of FAU 12 for the group with ASD when compared to the TD group. This appears to be an uncontrolled manifestation of smile expression. At the same time of this manifestation, the eye-tracker data reveal significantly averted gaze for the group with ASD. The group with ASD significantly avoids eye contact while interacting with a human facial image with expression. This aversion to

the face or facial image in conjunction with the higher prevalence of smile suggests a lack of congruency in the reciprocal communication such as social smile.

Finally, the second experiment involving dynamic facial expressions in response to dynamic audio-visual stimuli offers fine-grain computational models for evaluating differential traits in facial expressions. The second experiment suggests several metrics for computing the abnormality following several differential traits that are found with individuals with ASD. These differential traits in Chapter 7 suggest several behavioral biomarkers for ASD, which may be introduced in the screening procedures for the early detection of ASD-related traits. The presence of mouth frown (FAU 15), significantly lack of correlations in the activation of several FAU pairs, as well as the lack of FAU 6 (a constituent of a smile expression) demonstrate several measurable targets as differential traits that may be contributing to the oddity in the facial expressions of individuals with ASD. Interestingly, the first experiment also reveals a significantly differential trait in the prevalence of FAU 12 for the group with ASD. The differential traits found in the constituent FAUs of smile expressions (FAUs 6 and 12) in both experiments suggest that the smile expression may be a strong candidate for quantitative differential study of ASD. The second experiment also reveals flatness in the facial actions for the group with ASD in response to anxiety and anger of the peers. Thus, the emotional apathy or bluntness of the individual with ASD during a specific social context can be objectively computed from the facial response data.

Unlike the first experiment, a general trend in the second experiment shows that the TD group has produced higher facial activations when compared to the group with ASD. It appears that the difference in the stimuli type, tasks, and procedural contents of the experiments have played an important role in yielding variability in these results. Although the first experiment involves static 3D faces with expressions, the first experiment allows the subject to interact with the 3D face using a computer mouse. The task of recognizing and manipulating expressions from static 3D face may be trivial tasks for the TD group. However, interaction with computer-simulated objects may have created additional engagement and interests for the group with ASD as reported in prior studies [105, 106]. This may have led the group with ASD to create higher facial response than that of the TD group. This finding is consistent over multiple imaging modalities following the computational methods proposed in this dissertation (Chapters 5 and 6). The second experiment, however, demands the perception about the emotional content and context of

the dynamic audio-visual stimuli played by the animated Avatar character. The recognition of emotional content of such social stimuli may be more intellectually demanding for the subjects with ASD. Therefore, in the second experiment, the TD group has facially responded higher by recognizing the story content and context than the group with ASD. The flatness in the activation of FAUs for the group with ASD reveal their lower facial activations in Chapter 7.

While both experiments offer non-intrusive computational methods for investigating subtle facial responses, there are several advantages of the second experiment over the first one. The second experiment provides a dynamic and time-sampled facial response data unlike the static facial response captured by the 3D facial imaging sensor. A dynamic audio-visual stimulus is used in the second experiment, which is perhaps more effective than the static presentation of 3D faces with expression. Interestingly, the *Levator anguli oris* muscle region in the first experiment is at the similar location where mouth frown (FAU 15) takes place in the second experiment. Both of these targets located at a common facial region offer interesting differential traits for the group with ASD. Therefore, both experiments suggest a common target or differential trait for the screening of individuals with ASD. The asymmetric left and right facial response at *Levator anguli oris* muscle region and the unusual prevalence of mouth frown at the same facial region may be the measurable traits contributing to the pathophysiology of the disorder. Moreover, the facial motion capture system also offers a fine-grain analysis of a variety of facial muscle actions, which may not be feasible to probe using traditional electromyography (EMG) based electrodes. This is because EMG electrodes usually probe activations of only two facial muscles [33] and scanning all possible facial actions using intrusive electrodes may not be feasible. Furthermore, the proposed FAU-based method can detect subtle facial actions at a level that is beyond human visual perception.

8.1 Future Work

In this dissertation, the static and dynamic analyses of facial response in two consecutive experiments show several complementary and promising differential traits for the group with ASD. However, there are several limitations of the dissertation. Future studies are necessary to address these limitations for early detection and therapeutic intervention for ASD. Future studies are necessary in the following areas.

8.1.1 Full-scale study on larger population

One of the major limitations of the proposed pilot study is the smaller sample size. In the first experiment, there are only eight subjects in each group and the static acquisition of 3D facial imaging system yields only one facial image sample of the subject per trial. The dynamic acquisition and analysis of 3D facial data on a large population of subjects can further validate the observed differential traits for the group with ASD. For example, the asymmetry observed in the left and right facial muscle actions of the three individuals with ASD out of eight contributes to a statistically significant result for the group (See Section 5.5.2), which requires further study with larger population. Among the three individuals, two of them are siblings of 11 and 13 years of age who are expected to have a similar treatment history. The other individual is a 17-year old. The visual inspection of video images reveals that the 17-year old consistently showed smiles most likely because he had been reinforced so much to respond with a smile. Older children may have received training to respond and, hence, the bias in the differential traits due to the training can be difficult to identify. Therefore, it is important to focus on younger subjects who have not received significant training in regard to rendering facial expressions. On the other hand, the untrained younger subjects will help to identify if the differential traits in the response are genuinely because of the disorder. The study of younger subjects is also important for investigating early detection of behavioral markers.

8.1.2 Early detection of behavioral biomarkers

This dissertation identifies several differential traits in a group of school-aged children with ASD, which may be investigated even earlier during their preschool ages. The early detection of ASD-related differential traits is imperative for formulating subject-specific and effective intervention plans. In the future, an experimental paradigm involving parent-child or caregiver-child interaction may be planned during the early childhood period with appropriate visual, auditory, and gestural stimuli [128, 129]. An earlier detection of a trait related to a behavioral marker may help in early intervention for effective behavioral therapy and treatment. This will require facial expression datasets of younger children as offered by the Child Affective Face Set (CAFE) dataset [130]. The CAFE dataset provides a large number of ground truth facial expression imaging data of subjects of much younger age ranging from 2.7 to 8.7 years. These ground truth facial image data of control subjects from a much younger age group will offer an extremely useful reference to compare

similar affects of younger children with ASD. The availability of such early childhood facial expression datasets can play a crucial role in finding the early pathophysiology of ASD from the facial affect of much younger children. Consequently, this effort may require developing novel computational models for the early detection of ASD-related behavioral biomarkers.

8.1.3 Comparative studies to replace the use of EMG

One of the contributions of this dissertation includes non-intrusive computation of subtle facial responses instead of using intrusive sensors such as EMG. The non-intrusive imaging sensors also offers complete scanning of the face for detecting the subtle responses, which may not be feasible using EMG. Further studies will be required to compare the efficacy of facial imaging sensors with that of the EMG to study if the traditional EMG-based intrusive approach can be replaced with the application of the computation models proposed in this dissertation.

8.1.4 Adaptive learning system for behavioral engagement

The proposed behavioral biomarkers suggested in this dissertation offer valuable targets for future detection and intervention studies. A measurable trait in psychophysical response can be used as a metric to assess the efficacy of a training or intervention system for ASD therapeutics. This recommends the study and design of an adaptive intervention system where the stimuli are automatically adjusted by considering the quantitative information about the psychophysical response from the subjects. The differential traits obtained from the proposed computational models can be used to interactively drive an Avatar-based or humanoid-robot mediated or a hybrid adaptive learning systems for the therapy of subjects with ASD. Therefore, there is a need for careful investigation of adaptive and intelligent machines capable of collaborating and engaging subjects with ASD using quantitative and objective biomarker data. Such technology-assisted adaptive learning system may significantly reduce the requirement for time, cost, and training involving a human expert or caregiver for ASD.

BIBLIOGRAPHY

[1] J. Baio, "Prevalence of autism spectrum disorder among children aged 8 years," *Autism and Developmental Disabilities Monitoring Network, 11 Sites, USA* **63**, pp. 1–21, Mar. 2014.

[2] J. Baio, "Prevalence of autism spectrum disorders-autism and developmental disabilities monitoring network," *CDC: MMWR Surveillance Summaries* **61**(SS03), pp. 1–19, 2012.

[3] J. W. Harrington and K. Allen, "The clinician's guide to autism," *Pediatrics in review* **35**, pp. 62–78, Feb. 2014.

[4] S. Ozonoff, A. M. Iosif, F. Baguio, I. C. Cook, M. M. Hill, T. Hutman, S. J. Rogers, A. Rozga, S. Sangha, M. Sigman, M. B. Steinfeld, and G. S. Young, "A prospective study of the emergence of early behavioral signs of autism," *Journal of American Academy of Child and Adolescent Psychiatry* **49**, pp. 256–266, Mar 2010.

[5] W. Mandy, T. Charman, K. Puura, and D. Skuse, "Investigating the cross-cultural validity of DSM-5 autism spectrum disorder: Evidence from Finnish and UK samples," *Autism* **18**, pp. 45–54, Jan 2014.

[6] S. L. Lobar, "DSM-V Changes for Autism Spectrum Disorder (ASD): Implications for Diagnosis, Management, and Care Coordination for Children With ASDs," *Journal of Pediatric Health Care* , Oct 2015.

[7] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, Washington, DC, 2000.

[8] D. L. Robins, D. Fein, M. L. Barton, and J. A. Green, "The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders," *Journal of Autism Developmental Disorders* **31**, pp. 131–144, Apr 2001.

[9] M. C. Lai, M. V. Lombardo, B. Chakrabarti, and S. Baron-Cohen, "Subgrouping the autism "spectrum": Reflections on DSM-5," *PLOS Biolology* **11**(4), p. e1001544, 2013.

[10] S. Georgiades, "Investigating phenotypic heterogeneity in children with autism spectrum disorder: A factor mixture modeling approach," *Journal of Child Psychology and Psychiatry* **54**, pp. 206–215, Feb 2013.

[11] I. Roth, C. Barson, H. Rosa, G. Pasco, and T. Whatson, *The Autism Spectrum in the 21st Century: Exploring Psychology, Biology and Practice*, Jessica Kingsley Publications, Mar. 2010.

[12] J. A. Deisinger, *The Broad Autism Phenotype*, Emerald Group Publishing Limited., March 2015.

[13] J. Piven, P. Palmer, D. Jacobi, D. Childress, and S. Arndt, "Broader autism phenotype: evidence from a family history study of multiple-incidence autism families," *American Journal of Psychiatry* **154**, pp. 185–190, Feb 1997.

[14] G. Dawson, A. Estes, J. Munson, G. Schellenberg, R. Bernier, and R. Abbott, "Quantitative assessment of autism symptom-related traits in probands and parents: Broader Phenotype Autism Symptom Scale," *Journal of Autism and Developmental Disorders* **37**, pp. 523–536, Mar 2007.

[15] J. W. Harrington, R. Bai, and A. M. Perkins, "Screening children for autism in an urban clinic using an electronic M-CHAT," *Clinical Pediatrics* **52**, pp. 35–41, Jan 2013.

[16] D. J. Faso, N. J. Sasson, and A. E. Pinkham, "Evaluating posed and evoked facial expressions of emotion from adults with autism spectrum disorder," *Journal of Autism and Developmental Disorders* **45**, pp. 75–89, Jan 2015.

[17] M. A. Volker, C. Lopata, D. A. Smith, and M. L. Thomeer, "Facial encoding of children with high-functioning autism spectrum disorders," *Focus on Autism and other Development Disabilities* **24**, pp. 195–204, Oct 2009.

[18] D. Fein, *The neuropsychology of Autism*, Oxford University Press, April 2011.

[19] T. Guha, Z. Yang, A. Ramakrishna, R. B. Grossman, H. Darren, S. Lee, and S. S. Narayanan, "On Quantifying Facial Expression-Related Atypicality of Children with Autism Spectrum Disorder," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* **2015**, pp. 803–807, Apr 2015.

[20] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, "Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders," *Journal of Neuroscience Methods* **200**, pp. 237–256, Sep 2011.

[21] M. D. Samad and K. M. Iftekharuddin, "Frenet frame-based generalized space curve representation for pose-invariant classification and recognition of 3-D face," *IEEE Transactions on Human-Machine Systems* **PP**(99), pp. 1–12, 2016.

[22] M. D. Samad and K. M. Iftekharuddin, "Radial curve-based classification and localization of features for 3D facial expressions," *Proceedings of SPIE Three-Dimensional Imaging, Visualization, and Display* **9117**, June 2014.

[23] M. D. Samad, J. L. Bobzien, J. W. Harrington, and K. M. Iftekharuddin, "Non-intrusive optical imaging of face to probe physiological traits in autism spectrum disorder," *Optics and Laser Technology* **77**, pp. 221 – 228, 2016.

[24] M. D. Samad, J. L. Bobzien, J. W. Harrington, and K. M. Iftekharuddin, "Analysis of facial muscle activation in children with autism using 3D imaging," in *Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on*, pp. 337–342, Nov 2015.

[25] M. D. Samad, N. Diawara, J. Bobzien, J. Harrington, and K. M. Iftekharuddin, "Behavioral markers in facial and visual responses for autism spectrum disorders," *IEEE Transactions on Cognitive and Developmental Systems* (**Under Review**).

[26] M. D. Samad, J. Bobzien, J. Harrington, and K. M. Iftekharuddin, "Quantification of oddity in spontaneous facial actions of individuals with neurodevelopmental disorders," *IEEE Transactions on Affective Computing* (**Under Preparation**).

[27] F. Volkmar, C. Lord, B. A., S. R. T., and A. Klin, "Autism and pervasive developmental disorders," *Journal of Child Psychology and Psychiatry* **45**, pp. 135–170, Feb. 2004.

[28] B. Ruggeri, U. Sarkans, G. Schumann, and A. M. Persico, "Biomarkers in autism spectrum disorder: the old and the new," *Psychopharmacology* **231**, pp. 1201–1216, Mar 2014.

[29] W. Bosl, A. Tierney, H. Tager-Flusberg, and C. Nelson, "EEG complexity as a biomarker for autism spectrum disorder risk," *Biomedical Central Medicine* **9**, p. 18, 2011.

[30] G. Deshpande, L. Libero, K. R. Sreenivasan, H. Deshpande, and R. K. Kana, "Identification of neural connectivity signatures of autism using machine learning," *Frontiers in Human Neuroscience* **7**(670), 2013.

[31] A. T. Wang, S. S. Lee, M. Sigman, and M. Dapretto, "Reading affect in the face and voice: Neural correlates of interpreting communicative intent in children and adolescents with autism spectrum disorders," *Archives of General Psychiatry* **64**, pp. 698–708, Jun 2007.

[32] J. B. Wagner, S. B. Hirsch, V. K. Vogel-Farley, E. Redcay, and C. A. Nelson, "Eye-tracking, autonomic, and electrophysiological correlates of emotional face processing in adolescents with autism spectrum disorder," *Journal of Autism and Developmental Disorders* **43**, pp. 188–199, Jan 2013.

[33] C. Liu, K. Conn, N. Sarkar, and W. Stone, "Physiology-based affect recognition for computer-assisted intervention of children with autism spectrum disorder," *International Journal of Human-Computer Studies* **66**(9), pp. 662 – 677, 2008.

[34] K. Aldridge, I. D. George, K. K. Cole, J. R. Austin, T. N. Takahashi, Y. Duan, and J. H. Miles, "Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes," *Molecular Autism* **2**(1), p. 15, 2011.

[35] H. Ozgen, G. S. Hellemann, M. V. de Jonge, F. A. Beemer, and H. van Engeland, "Predictive value of morphological features in patients with autism versus normal controls," *Journal of Autism and Developmental Disorders* **43**, pp. 147–155, Jan 2013.

[36] P. Hammond, C. Forster-Gibson, A. E. Chudley, J. E. Allanson, T. J. Hutton, S. A. Farrell, J. McKenzie, J. J. Holden, and M. E. Lewis, "Face-brain asymmetry in autism spectrum disorders," *Molecular Psychiatry* **13**, pp. 614–623, Jun 2008.

[37] K. Rice, J. M. Moriuchi, W. Jones, and A. Klin, "Parsing heterogeneity in autism spectrum disorders: visual scanning of dynamic social scenes in school-aged children," *Journal of American Academy of Child Adolescence and Psychiatry* **51**, pp. 238–248, Mar 2012.

[38] T. Dratsch, C. Schwartz, K. Yanev, L. Schilbach, K. Vogeley, and G. Bente, "Getting a grip on social gaze: control over others' gaze helps gaze detection in high-functioning autism," *Journal of Autism and Developmental Disorders* **43**, pp. 286–300, Feb 2013.

[39] E. M. Thabet and H. S. Zaghloul, "Auditory profile and high resolution CT scan in autism spectrum disorders children with auditory hypersensitivity," *European Archive of Otorhinolaryngology* **270**, pp. 2353–2358, Aug 2013.

[40] U. Lahiri, E. Bekele, E. Dohrmann, Z. Warren, and N. Sarkar, "Design of a virtual reality based adaptive response technology for children with autism," *IEEE Transaction on Neural Systems and Rehabilitation Engineering* **21**, pp. 55–64, Jan 2013.

[41] J. J. Thompson, *Beyond Words: Nonverbal Communication in the Classroom.*, MacMillan Publishing Company, New York, April 1973.

[42] S. Nowicki and M. Duke, *Helping the Child Who Doesnt Fit In.*, Peachtree Publishers, Atlanta, GA, January 1992.

[43] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, Washington, DC, 2000.

[44] K. M. Rump, J. L. Giovannelli, N. J. Minshew, and M. S. Strauss, "The development of emotion recognition in individuals with autism," *Child Development* **80**(5), pp. 1434–1447, 2009.

[45] R. B. Grossman and H. Tager-Flusberg, "Reading Faces for Information about Words and Emotions in Adolescents with Autism," *Research on Autism Spectrum Disorders* **2**, pp. 681–695, Oct 2008.

[46] P. M. Beall, E. J. Moody, D. N. McIntosh, S. L. Hepburn, and C. L. Reed, "Rapid facial reactions to emotional facial expressions in typically developing children and children with autism spectrum disorder," *Journal of Experimental Child Psychology* **101**, pp. 206–223, Nov 2008.

[47] M. Hoque, D. McDuff, and R. Picard, "Exploring temporal patterns in classifying frustrated and delighted smiles," *Affective Computing, IEEE Transactions on* **3**, pp. 323–334, July 2012.

[48] P. L. Dhingra and S. Dhingra, *Diseases of Ear, Nose and Throat.*, Elsevier India, December 2013.

[49] M. Dapretto, M. S. Davies, J. H. Pfeifer, A. A. Scott, M. Sigman, S. Y. Bookheimer, and M. Iacoboni, "Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders," *Nature Neuroscience* **9**, pp. 28–30, Jan 2006.

[50] E. Pellicano, G. Rhodes, and A. J. Calder, "Reduced gaze aftereffects are related to difficulties categorising gaze direction in children with autism," *Neuropsychologia* **51**, pp. 1504–1509, Jul 2013.

[51] S. R. Zaki and S. A. Johnson, "The role of gaze direction in face memory in autism spectrum disorder," *Autism Research* **6**, pp. 280–287, Aug 2013.

[52] Prelock, P. A., *Autism Spectrum Disorders: Issues in Assessment and Intervention.*, Pro-ED Inc.; 1st edition, 2006.

[53] D. Mathersul, S. McDonald, and J. A. Rushby, "Automatic facial responses to affective stimuli in high-functioning adults with autism spectrum disorder," *Physiology and Behavior* **109**, pp. 14–22, Jan 2013.

[54] A. Rozga, T. Z. King, R. W. Vuduc, and D. L. Robins, "Undifferentiated facial electromyography responses to dynamic, audio-visual emotion displays in individuals with autism spectrum disorders," *Developmental Science* **16**, pp. 499–514, Jul 2013.

[55] J. L. Adrien, A. Perrot, D. Sauvage, I. Leddet, C. Larmande, L. Hameury, and C. Barthelemy, "Early symptoms in autism from family home movies. Evaluation and comparison between 1st and 2nd year of life using I.B.S.E. scale," *Acta paedopsychiatrica* **55**(2), pp. 71–75, 1992.

[56] M. Stel, C. van den Heuvel, and R. C. Smeets, "Facial feedback mechanisms in autistic spectrum disorders," *Journal of Autism and Devevelopmental Disorders* **38**, pp. 1250–1258, Aug 2008.

[57] G. Simons, M. C. Pasqualini, V. Reddy, and J. Wood, "Emotional and nonemotional facial expressions in people with Parkinson's disease," *Journal of the International Neuropsychological Society* **10**, pp. 521–535, Jul 2004.

[58] M. Tomasello, "Do apes ape?," *C. Heyes & B. Galef (Eds.) Social Learning in Animals: The roots of Culture* , pp. 319–346, 1996.

[59] S. Park and K. Kim, "Physiological reactivity and facial expression to emotion-inducing films in patients with schizophrenia," *Archives of Psychiatric Nursing* **25**, pp. 37–47, Dec 2011.

[60] L. M. Oberman, P. Winkielman, and V. S. Ramachandran, "Slow echo: Facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders," *Developmental Science* **12**, pp. 510–520, Jul 2009.

[61] M. J. Magnee, B. de Gelder, H. van Engeland, and C. Kemner, "Facial electromyographic responses to emotional information from faces and voices in individuals with pervasive developmental disorder," *Journal of Child Psychology and Psychiatry* **48**, pp. 1122–1130, Nov 2007.

[62] J. Hashemi, M. Tepper, T. Vallin Spina, A. Esler, V. Morellas, N. Papanikolopoulos, H. Egger, G. Dawson, and G. Sapiro, "Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants," *Autism Research and Treatment* **2014**, p. 935686, 2014.

[63] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," in *Automatic Face and Gesture Recognition, Proceedings of Third IEEE International Conference on*, pp. 336–341, Apr 1998.

[64] P. S. Aleksic and A. K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream hmms," *IEEE Transaction on Information Forensics and Security* **1**, pp. 3–11, March 2006.

[65] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pp. 3444–3451, June 2013.

[66] L. Ding, X. Ding, and C. Fang, "Continuous pose normalization for pose-robust face recognition," *IEEE Signal Processing Letters* **19**, pp. 721–724, Nov 2012.

[67] X. X. Li, D. Q. Dai, X. F. Zhang, and C. X. Ren, "Structured sparse error coding for face recognition with occlusion," *IEEE Transaction on Image Processing* **22**, pp. 1889–1900, May 2013.

[68] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang, "Total variation models for variable lighting face recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence* **28**, pp. 1519–1524, Sept 2006.

[69] X. Xie, W. S. Zheng, J. Lai, P. C. Yuen, and C. Y. Suen, "Normalization of face illumination based on large-and small-scale features," *IEEE Transaction on Image Processing* **20**, pp. 1807–1821, July 2011.

[70] U. Prabhu, J. Heo, and M. Savvides, "Unconstrained pose-invariant face recognition using 3D generic elastic models," *IEEE Transaction on Pattern Analysis and Machine Intelligence* **33**, pp. 1952–1961, October 2011.

[71] H. Mohammadzade and D. Hatzinakos, "Iterative closest normal point for 3D face recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence* **35**(2), pp. 381–397, 2013.

[72] Y. Lei, M. Bennamoun, and A. A. El-Sallam, "An efficient 3D face recognition approach based on the fusion of novel local low-level features," *Pattern Recognition* **46**, pp. 24–37, 2013.

[73] A. Maalej, B. B. Amor, M. Daoudi, A. Srivastava, and S. Berretti, "Shape analysis of local facial patches for 3D facial expression recognition," *Pattern Recognition* **44**, pp. 1581–1589, 2011.

[74] A. Savran, B. Sankur, and M. T. Bilge, "Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units," *Pattern Recognition* **45**, pp. 767–782, 2012.

[75] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, "Recognition of 3D facial expression dynamics," *Image and Vision Computing* **30**, pp. 762–773, 2012.

[76] J. Wang, L. Yin, X. Wei, and Y. Sun, "3D facial expression recognition based on primitive surface feature distribution," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* **2**, pp. 1399–1406, 2006.

[77] H. Li, J.-M. Morvan, and L. Chen, "3D facial expression recognition based on histograms of surface differential quantities," *Proceedings of the 13th International Conference on Advanced Concepts for Intelligent Vision Systems* , pp. 483–494, 2011.

[78] J. Sung and D. Kim, "Pose-robust facial expression recognition using view-based 2D+3D AAM," *IEEE Transaction on System, Man, and Cybernetics.* **38**, pp. 852–866, July 2008.

[79] B. Gong, Y. Wang, J. Liu, and X. Tang, "Automatic facial expression recognition on a single 3D face by exploring shape deformation," *Proceedings of the 17th ACM International Conference on Multimedia* , pp. 569– 572, 2009.

[80] M. Drira, B. Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D face recognition under expressions occlusions and pose variations," *IEEE Transaction on Pattern Analysis and Machine Intelligence* , 2013.

[81] D. Smeets, J. Hermans, D. Vandermeulen, and P. Suetens, "Isometric deformation invariant 3D shape recognition," *Pattern Recognition* **45**, pp. 2817–2831, 2012.

[82] H. Tang, B. Yin, Y. Sun, and Y. Hu, "3D face recognition using local binary patterns," *Signal Processing* **93**, pp. 2190–2198, August 2013.

[83] X. Li and F. Da, "Efficient 3D face recognition handling facial expression and hair occlusion," *Image and Vision Computing* **30**, pp. 668–679, Sept. 2012.

[84] Q. Ju, "Robust binary neural networks based 3D face detection and accurate face registration," *International Journal of Computational Intelligence Systems* **6**, pp. 669–683, July 2013.

[85] Sangineto, "Pose and expression independent facial landmark localization using dense-surf and the hausdorff distance," *IEEE Transaction on Pattern Analysis and Machine Intelligence* **35**, pp. 624–638, March 2013.

[86] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression invariant representations of faces," *IEEE Transaction Image Processing* **16**(1), pp. 188–197, 2007.

[87] P. Liu, Y. Wang, D. Huang, Z. Zhang, and L. Chen, "Learning the spherical harmonic features for 3D face recognition," *IEEE Transaction on Image processing* **22**, pp. 914–924, March 2013.

[88] S. Jahanbin, R. Jahanbin, and A. C. Bovik, "Passive three dimensional face recognition using iso-geodesic contours and procrustes analysis," *International Journal of Computer Vision* **105**, pp. 87–108, June 2013.

[89] P. Ekman and W. Friesen, "Facial Action Coding System (FACS): A technique for the measurement of facial action, Palo Alto, CA: Consulting," 1978.

[90] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Classifying facial actions," *IEEE Transaction Pattern Analysis and Machine Intelligence* **21**, pp. 974–989, Oct 1999.

[91] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," *Proceedings of National Academy of Science, U.S.A.* **111**, pp. E1454– E1462, Apr 2014.

[92] T. Kanade, J. Cohn, and Y. L. Tian, "Comprehensive database for facial expression analysis," *IEEE International Conference Automatic Face and Gesture Recognition* , p. 46, 2000.

[93] J. Cohn, T. Kruez, I. Matthews, Y. Yang, M. H. Nguyen, M. Padilla, F. Zhou, and F. De la Torre, "Detecting depression from facial actions and vocal prosody," in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pp. 1–7, Sept 2009.

[94] T. Wu, N. J. Butko, P. Ruvolo, J. Whitehill, M. S. Bartlett, and J. R. Movellan, "Multilayer Architectures for Facial Action Unit Recognition," *IEEE Transaction on Systems, Man, and Cybernetics - B Cybernetics* , May 2012.

[95] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, "An efficient 3D face recognition approach using local geometrical signatures," *Pattern Recognition* **47**(2), pp. 509–524, 2014.

[96] O. Ocegueda, T. Fang, S. K. Shah, and I. A. Kakadiaris, "3D face discriminant analysis using gauss-markov posterior marginals," *IEEE transaction on Pattern Analysis and Machine Intelligence* **35**, pp. 728–739, March 2013.

[97] Y. Wang and X. Ma, "3D facial expression recognition based on encoded templates," *Symposium on Photonics and Optoelectronics* , pp. 1–4, 2012.

[98] A. E. Youssef, S. F. Aly, A. S. Ibrahim, and A. L. Abbott, "Auto-optimized multi-modal expression recognition framework using 3D kinect data for ASD therapeutic aid," *International Journal of Modeling and Optimization* **3**, pp. 112–115, April 2013.

[99] C. Samir, A. Srivastava, M. Daoudi, and E. Klassen, "An intrinsic framework for analysis of facial surfaces," *International Journal of Computer Vision* **82**, pp. 80–95, April 2009.

[100] M. D. Lerner, J. C. McPartland, and J. P. Morris, "Multimodal emotion processing in autism spectrum disorders: An event-related potential study," *Developmental Cognitive Neuroscience* **3**, pp. 11–21, Jan 2013.

[101] K. Edwards, "The face of time: Temporal cues in facial expressions of emotion," *Psychological Science* **9**, pp. 270–276, 1998.

[102] Z. C. Chen, M. N. Albdour, J. A. Lizardo, Y. A. Chen, and P. K. Chen, "Precision of three-dimensional stereo-photogrammetry (3DMD) in anthropometry of the auricle and its application in microtia reconstruction," *Journal of Plastic, Reconstructive and Aesthetic Surgery* **68**, pp. 622–631, May 2015.

[103] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3D dynamic facial expression database," *Proceedings of IEEE Int'l Conference Automatic Face and Gesture Recognition* , pp. 1–6, Sept. 2008.

[104] D. J. Moore, P. McGrath, and J. Thrope, "Computer aided learning for people with autism a framework for research and development.," *Innovation in Education and Training International* **37**(3), pp. 218–228, 2000.

[105] B. O. Ploog, A. Scharf, D. Nelson, and P. J. Brooks, "Use of computer-assisted technologies (CAT) to enhance social, communicative, and language development in children with autism spectrum disorders," *Journal of Autism and Developmental Disorders* **43**, pp. 301–322, Feb 2013.

[106] I. M. Hopkins, M. W. Gower, T. A. Perez, D. S. Smith, F. R. Amthor, F. C. Wimsatt, and F. J. Biasini, "Avatar assistant: Improving social skills in students with an

ASD through a computer-based intervention," *Journal of Autism and Developmental Disorders* **41**, pp. 1543–1555, Nov 2011.

[107] A. Gray, *Modern Differential Geometry of Curves and Surfaces*, CRC Press Inc., Florida, USA, 2000.

[108] Boehm, W. and Prautzsch, *Geometric Concepts for Geometric Design*, A K Peters Wellesley, Massachusetts, 1994.

[109] S. Zucker, "Differential geometry from the frenet point of view: Boundary detection, stereo, texture and color," *Handbook in Mathematical Models in Computer Vision* , pp. 357–373, 2006.

[110] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, M. Jaesik, and W. Worek, "Overview of the face recognition grand challenge.," *Computer Vision and Pattern Recognition (CVPR)* **1**, pp. 947–954, June 2005.

[111] A. Savran and L. Akarun, "Bosphorous database for 3D face analysis," *Proceedings of first COST 2101 Workshop Biometrics and Identity Management* , pp. 47–56, May 2008.

[112] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transaction on Pattern Analysis and Machine intelligence* **27**(8), pp. 1226–1238, 2005.

[113] N. Alyuz, B. Gokberk, and L. Akarun, "Regional registration and curvature descriptors for expression resistant 3D face recognition," *Proceedings of 17th IEEE Signal Processing and Communication Applied Conference* , pp. 544–547, Apr. 2009.

[114] M. Husken, M. Brauckmann, S. Gehlen, and C. V. d. Malsburg, "Strategies and benefits of fusion of 2D and 3D face recognition," *Proceedings of IEEE Workshop Face Recognition Grand Challenge Experiments* , pp. 174–181, June 2005.

[115] F. AL-Osaimi, M. Bennamoun, and A. Mian, "An expression deformation approach to non-rigid 3D face recognition," *International Journal of Computer Vision* **81**, pp. 302–316, Mar. 2009.

[116] T. Faltemier, K. Bowyer, and P. Flynn, "A region ensemble for 3D face recognition," *IEEE Transaction on Information Forensics and Security* **3**, pp. 62–73, Mar. 2008.

[117] N. Vretos, N. Nikolaidis, and I. Pitas, "3D facial expression recognition using zernike moments on depth images," *18th IEEE International Conference on Image Processing* , pp. 773–776, 2011.

[118] Y. V. Venkatesh, A. A. Kassim, J. Yuan, and T. D. Nguyen, "On the simultaneous recognition of identity and expression from bu-3Dfe datasets.," *Pattern Recognition Letters* **33**, pp. 1785–1793, 2012.

[119] C. Zhong, T. Sun, and T. Tan, "Robust 3D face recognition using learned visual codebook," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* , pp. 1–6, 2007.

[120] X. Zhu and D. Ramanan, "Face detection, pose estimation and landmark localization in the wild," *Computer Vision and Pattern Recognition (CVPR) Providence* , June 2012.

[121] W. Graf and H. Krueger, "Ergonomic evaluation of user-interfaces by means of eye-movement data," *Proceedings of the 3rd International Conference on Human-Computer Interaction* , pp. 659–665, 1989.

[122] K. Humphreys, N. Minshew, G. L. Leonard, and M. Behrmann, "A fine-grained analysis of facial expression processing in high-functioning adults with autism," *Neuropsychologia* **45**, pp. 685–695, Mar 2007.

[123] H. Oster, "Emotion in the infant's face: Insights from the study of infants with facial anomalies," *Annals of New York Academy of Sciences* **1000**, pp. 197–204, Dec 2003.

[124] L. Capps, C. Kasari, N. Yirmiya, and M. Sigman, "Parental perception of emotional expressiveness in children with autism," *Journal of Consulting and Clinical Psychology* **61**, pp. 475–484, Jun 1993.

[125] R. N. Emde, T. J. Gaensbauer, and R. J. Harmon, "Emotional expression in infancy: A biobehavioral study," *Psychology Issues* **10**(01), pp. 1–200, 1976.

[126] R. B. Grossman, L. R. Edelson, and H. Tager-Flusberg, "Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism," *Journal of Speech Language and Hearing Research* **56**, pp. 1035–1044, Jun 2013.

[127] B. Atchison and D. Dirette, *Conditions in Occupational Therapy: Effect on Occupational Performance*, Lippincott Williams and Wilkins, November 2011.

[128] K. Holmboe, M. Elsabbagh, A. Volein, L. A. Tucker, S. Baron-Cohen, P. Bolton, T. Charman, and M. H. Johnson, "Frontal cortex functioning in the infant broader autism phenotype," *Infant Behavior and Development* **33**, pp. 482–491, Dec 2010.

[129] S. Lloyd-Fox, A. Blasi, C. E. Elwell, T. Charman, D. Murphy, and M. H. Johnson, "Reduced neural sensitivity to social stimuli in infants at risk for autism," *Proceedings of Biological Sciences* **280**, p. 20123026, May 2013.

[130] V. LoBue and C. Thrasher, "The child affective facial expression (CAFE) set: Validity and reliability from untrained adults," *Frontiers in Emotion Science* **5**.

APPENDIX A

IRB APPROVAL

No.: 13-167

OLD DOMINION UNIVERSITY
HUMAN SUBJECTS INSTITUTIONAL REVIEW BOARD
RESEARCH PROPOSAL REVIEW NOTIFICATION FORM

TO: Khan Iftekharuddin   DATE: **October 17, 2013**
Responsible Project Investigator   *IRB Decision Date*

3-D Facial Expression and Visual Stimuli Analysis for Autism Spectrum Disorders (ASD)
Name of Project

Please be informed that your research protocol has received approval by the Institutional
Review Board. Your research protocol is:

    __ Approved
    ___ Tabled/Disapproved
    _X_ Approved, (Progress Report) contingent on making the changes below*

_____   **October 17, 2013**
    IRB Chairperson's Signature   *date*

Contact the IRB for clarification of the terms of your research, or if you wish to make
ANY change to your research protocol.

The approval expires one year from the IRB decision date. You must submit a Progress
Report and seek re-approval if you wish to continue data collection or analysis beyond
that date, or a Close-out report. You must report adverse events experienced by subjects
to the IRB chair in a timely manner (see university policy).

---

*   Approval of your research is CONTINGENT upon the satisfactory completion of
the following changes and attestation to those changes by the chairperson of the
Institutional Review Board. Research may not begin until after this attestation.
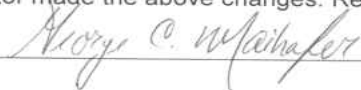
*In the Progress Report
    • On the title page, the end date needs to be changed to be that date that
    the report was submitted to the Office of Research (09/26/2013).
In the Consent Form
Submit the new consent form that states the revised age group of the participants.

Attestation
As directed by the Institutional Review Board, the Responsible Project
Investigator made the above changes. Research may begin.

_____   **December 18, 2013**

APPENDIX B

INCLUSIONARY/EXCLUSIONARY CRITERIA

An eligible subject should comply with the inclusion criteria as below. The eligible subject should also not have the issues listed in the exclusion criteria below that would keep him/her from participating in this study.

- Inclusion criteria Adolescent and adult (ages 7 to 20) male and female individuals, diagnosed with an Autism Spectrum Disorder by documentation provided by the subject, IQ score > 70, and the ability to comprehend the verbal content of the psychological tests being administered as determined by conversations during the phone screens and consent procedure.

- Exclusion criteria individuals, not diagnosed with Autism Spectrum Disorder, not able to comprehend the verbal content of the psychological tests being administered including the following:

    - Heart, lung, kidney, or liver disease, or if your child is on dialysis

    - High blood pressure diagnosis or hypertension

    - Cancer diagnosis or tumors of the brain

    - Neurological illnesses (e.g. multiple sclerosis) or history of concussion/head injury/seizures

APPENDIX C

SCRIPT FOR THE AUDIO-VISUAL STIMULUS

Hi, my name is Robert. Glad to have you here today.

How are you? (pause)

(hmmI see! Alright!)

I am in 8th grade in Blair middle schoolDo you go to school?...(pause)... Which grade are you in? (Pause)

Hmm I see !

The thing I like about being at school is that I have many wonderful friendsthe teachers are greatbut sometimes I would rather play baseball outside than do my homework! Also, in my free time at home, I like to play mine craft. What is your favorite game to play?

Last week, my friends and I went to see the Batman movie. Next week we are going to watch Transformer 4, it is going to be awesome!

Friday night my friends and I like to go to the pizzeria.I like to get pepperoni, however, Tony likes anchovies on his slice. Do you like anchovies?

Hmm I see.

You know what happened last week?(pause)

Tony was sick on Monday and Tuesday with a cold, so I let Tony borrow my math notebook last Wednesday. He said he would return by Thursday, but he forgot to bring it to school! We have a test tomorrow. My friend Sarah said I could copy hers, but I can't read her handwriting!

Now, I am on the bus home and I don't have my notes!! I'm really angry with Tony. How would you feel to be in my situation?(Pause...)

I am very worried that I will not be ready for Mr. Connery Math exam. Sarah says everything will be fine. She invited me to her birthday party at the bowling alley Saturday. Here is my bus stop.

I told my mom that I don't have my notes. She is calling Tony's dad. Oh good, Tony's dad is bringing my notes to my house.

Tony also brought me a Snickers bar! Wow! He remembered that it was my favorite candy bar. I am so relieved to have my notes and an extra treat.

Here I am back in school. Mr. Connery is about to give me my exam. I feel like I'm prepared.

I am in my exam. At first the exam, seemed tricky, but when I stopped to collect my thoughts I was able to figure out 3 of the 4 questions. Do you sometimes feel overwhelmed during tests? (Pause)

It Saturday! I am at the bowling party with 15 of Sarah's friends. We are having a great time. There is plenty of pizza for everyone.

I started off not bowling so well. I put the ball in the gutter three times this game. I'm still having fun though. This is last my bowl. I knocked all of the pins over! Wow the only strike of this game! Now I get another chance. Here I go woah I knocked down 9 pins and the last 1 is wobbling. Come on pin, fall over!. If I hit 5 pins over I will have the highest score. Here it goes Whoops the ball went left and only hit 4 pins over. Oh well, second place is still good!

Thanks for listening for my week.

VITA

Manar D. Samad

Department of Electrical and Computer Engineering

Old Dominion University

Norfolk, VA 23529

Manar D. Samad received his M.S. degree in Computer Engineering from the University of Calgary in 2011 and the B.S. degree in Electrical and Electronic Engineering from the Bangladesh University of Engineering and Technology in 2007. As a graduate research assistant, he worked under the supervision of neuroArm, one of the pioneers of the robot-assisted neurosurgical system. There, he studied the effect of force feedback and 3D stereoscopic vision on the perception and performance of robot-assisted surgical tasks. In Spring 2012, he started his Ph.D. program in the ECE Vision Lab at Old Dominion University. The motivation of his current research is to provide quantitative and non-intrusive methods for sensing nonverbal behavior of a patient, which are currently qualitative and intrusive in the clinical practice. Prior to his graduate studies, he taught electrical engineering courses at the undergraduate level for two years at the United International University in Dhaka, Bangladesh. His research interests include Computer Vision, Machine Learning, Affective Computing, Virtual Reality, and Human-Computer Interactions. Below are some of his selected publications as first author.

- **M. Samad** and K. Iftekharuddin, Frenet frame-based generalized space curve representation for pose invariant classification and recognition of 3D Face. *IEEE Transaction on Human-Machine Systems (THMS)*, vol. pp, no. 99, pp. 1-12, doi=10.1109/THMS.2016.2515602.

- **M. Samad**, J. Bobzien, J. Harrington, and K. Iftekharuddin, Non-intrusive optical imaging of face to probe physiological traits in autism spectrum disorder. *Journal of Optics and Laser Technology* Vol. 77, pp. 221-228, Mar. 2016.

- **M. Samad**, J. Bobzien, J. Harrington, and K. Iftekharuddin, Analysis of facial muscle actions in children with autism using 3D imaging *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 337-342, Washington DC, USA, Nov. 2015.