6-26-2019

Non-Intrusive Affective Assessment in the Circumplex Model from Pupil Diameter and Facial Expression Monitoring

Sudarat Tangnimitchok
Florida International University, stang018@fiu.edu

Recommended Citation

Tangnimitchok, Sudarat, "Non-Intrusive Affective Assessment in the Circumplex Model from Pupil Diameter and Facial Expression Monitoring" (2019). *FIU Electronic Theses and Dissertations*. 4224.
https://digitalcommons.fiu.edu/etd/4224

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

NON-INTRUSIVE AFFECTIVE ASSESSMENT IN THE CIRCUMPLEX MODEL FROM PUPIL DIAMETER AND FACIAL EXPRESSION MONITORING

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

by

Sudarat Tangnimitchok

2019

To: Dean John L. Volakis
College of Engineering and Computing

This dissertation, written by Sudarat Tangnimitchok, and entitled Non-Intrusive Affective Assessment in the Circumplex Model from Pupil Diameter and Facial Expression Monitoring, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Malek Adjouadi

Jean H. Andrian

Ruogu Fang

Armando Barreto, Major Professor

Date of Defense: June 26, 2019

The dissertation of Sudarat Tangnimitchok is approved.

Dean John L. Volakis
College of Engineering and Computing

Andres G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2019

DEDICATION

To my family, my beloved friends and people I have met during my journey.

ACKNOWLEDGMENTS

To my mom who is always supportive and worried about me every single day.

To my advisor who is always understanding and push me in the right direction. He gives an honest and dedicated advise every single time I lost my way.

To my friend Moon who is together with me in the loneliest phase of my life.

ABSTRACT OF THE DISSERTATION

NON-INTRUSIVE AFFECTIVE ASSESSMENT IN THE CIRCUMPLEX MODEL FROM PUPIL DIAMETER AND FACIAL EXPRESSION MONITORING

by

Sudarat Tangnimitchok

Florida International University, 2019

Miami, Florida

Professor Armando Barreto, Major Professor

Automatic methods for affective assessment seek to enable computer systems to recognize the affective state of their users. This dissertation proposes a system that uses non-intrusive measurements of the users pupil diameter and facial expression to characterize his /her affective state in the Circumplex Model of Affect. This affective characterization is achieved by estimating the affective arousal and valence of the users affective state.

In the proposed system the pupil diameter signal is obtained from a desktop eye gaze tracker, while the face expression components, called Facial Animation Parameters (FAPs) are obtained from a Microsoft Kinect module, which also captures the face surface as a cloud of points. Both types of data are recorded 10 times per second. This dissertation implemented pre-processing methods and fixture extraction approaches that yield a reduced number of features representative of discrete 10-second recordings, to estimate the level of affective arousal and the type of affective valence experienced by the user in those intervals.

The dissertation uses a machine learning approach, specifically Support Vector Machines (SVMs), to act as a model that will yield estimations of valence and arousal from the features derived from the data recorded.

Pupil diameter and facial expression recordings were collected from 50 subjects who volunteered to participate in an FIU IRB-approved experiment to capture their reactions to the presentation of 70 pictures from the International Affective Picture System (IAPS) database, which have been used in large calibration studies and therefore have associated arousal and valence mean values. Additionally, each of the 50 volunteers in the data collection experiment provided their own subjective assessment of the levels of arousal and valence elicited in him / her by each picture. This process resulted in a set of face and pupil data records, along with the expected reaction levels of arousal and valence, i.e., the labels, for the data used to train and test the SVM classifiers.

The trained SVM classifiers achieved 75% accuracy for valence estimation and 92% accuracy in arousal estimation, confirming the initial viability of non-intrusive affective assessment systems based on pupil diameter and face expression monitoring.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF FIGURES

FIGURE PAGE

CHAPTER 1

INTRODUCTION

1.1 Motivation

Affective Computing was first introduced by Rosalind Picard [P$^+$95] in 1997. It proposes that interactions between humans and computers can take place at an affective level. To be more specific, it aims to enable a computer to understand its user's emotion and be able to respond appropriately or even sympathize with its user's affective state. Some might challenge the benefit of such goal, pointing out that maybe it is not necessary for the computer to have emotional abilities since a computer is just a tool and it is fine to keep it as a rigid tool. The argument is logical and reasonable; however, there are some situations where human-computer interaction at an affective level can improve the user's experience significantly by adding the user's emotional information, such as frustration, interest, displeasure, and etc., to the process implemented in the computer so it can respond in an appropriate way. Here are some of the applications in which we can apply affective computing to enhance a user's experience.

- **Lessen the User's Frustration**

  Many users, over time, show a lot of frustration toward computers. ”A widely-publicized 1999 study by Concord Communications in the U.S. found that 84% of help-desk managers surveyed said that users admitted to engaging in violent and abusive behavior toward computers” (Quote from [Pic99]). This fact is one of the reasons why Human-Computer Interaction (HCI) researchers strive to lessen users' frustration during their interaction with a computer via the interface design but, unfortunately, the frustration is bound to happen in some way or another. As an alternative way to deal with the user's frustration, computers should learn how to lessen the user's frustration or displeasure. An example to reduce the user's frustration could make the system play relaxing music when it detects some certain threshold of stress from its user.

- **Online-Based Education**

  E-Learning is an innovative way of learning via electronic resources, typically on the internet. Students can choose freely what contents to consume completely at their own pace and time. Due to the flexibility that E-learning provides, it has become increasingly popular as time passed. However, Online courses have one big disadvantage, which is the lack of interaction between the teacher and the student, because it is difficult for the teacher to properly monitor his/her student reactions when the class is conducted remotely. Hence, a good example of how affective computing system can be useful in the addition of affective abilities to the online classroom system, where a computer can monitor the level of student's engagement and stress during the lectures, especially with younger students.

- **Online-Based Services**

  Online-Based services will undoubtedly be used by service providers in the future. We have seen many service providers start to integrate online-based services to give more flexible options to their customers. For instance, some health providers now offer E-therapy, i.e., online-based health consultations, to patients so they do not have to travel to a hospital in person, or in an emergency case, the health provider can provide the advice right away, in real-time. Another relevant example is the increasing use of online customer service. Most recently, a new practice that many big companies have adopted is the use of customer Service Bots, which automate their customer service with AI chatbots [1] to solve a simple routine problem that does not involve complicated tasks that could require human intelligence. This way of dealing with customers provides resilience capacities for companies to deal with the situation when a lot of customers phone in at the same time. Even though an AI chatbot is a very efficient way to provide customer support, a robot is simply a robot. Currently, robots cannot interact at an emotional level with customers. Incorporating affective abilities to the chatbot will enhance customer experience significantly. Besides, companies are also interested in collecting data of customer feedback so an ability to detect the customer's satisfaction during the service will be highly valuable to companies for improving their services.

- **Assistive Technology**

  Individuals suffering from autism who tend to have a social-emotional communicative impairment that makes it difficult to interact with other people. Using computers or assisting technology to communicate with non-autistics may help in easing this difficulty by allowing an autistic person to communicate non-verbally with others. Affective technology can help autistics to identify non-autistics' affective states which are often difficult for the autistic person. Additionally, current intervention techniques suggest that intensive and progressive training can help autistics to improve their social-emotional

[1] Artificially Intelligent chatbot

skills and in recognizing other people's emotion. That is why affective technology can help to assist autistics to develop their social-emotional capabilities.

Human has strived to develop intelligent systems. However, most of the time, the emotional aspects of intelligence are ignored, as they are seen as less critical. However, this topic is also very important to balance the way humans interact with computers. HCI researchers should always keep this thought in mind while researching a better way of enhancing a user's experience with computers.

1.2 Affective Computing

The idea of enabling a computer to generate empathy and be able to be empathetic to its user is a very challenging goal. The difficulties associated with the actual implementation of an affective computing system might be best appreciated if one considers the 3 fundamental tasks that must be performed to fully animate the performance of an affective computing system (affective computer), as outlined by Hudlicka [Hud03]: These tasks can be described as (See Figure 1.1):

1. Affect Sensing and Recognition
2. User Affect Modeling / Machine Affect Modeling
3. Machine Affect Expression

The affective sensing and recognition tasks aim at making the machine aware of the affective state of the human user. This will require sensing some observable manifestations of that affective status and recognizing (or cataloging) the state, so that, then, the machine may determine (by following some pre-programmed interplay guidelines) which affective state it should adopt in response, and, further, the type of affective expression that it should present to the user. Those initial stages

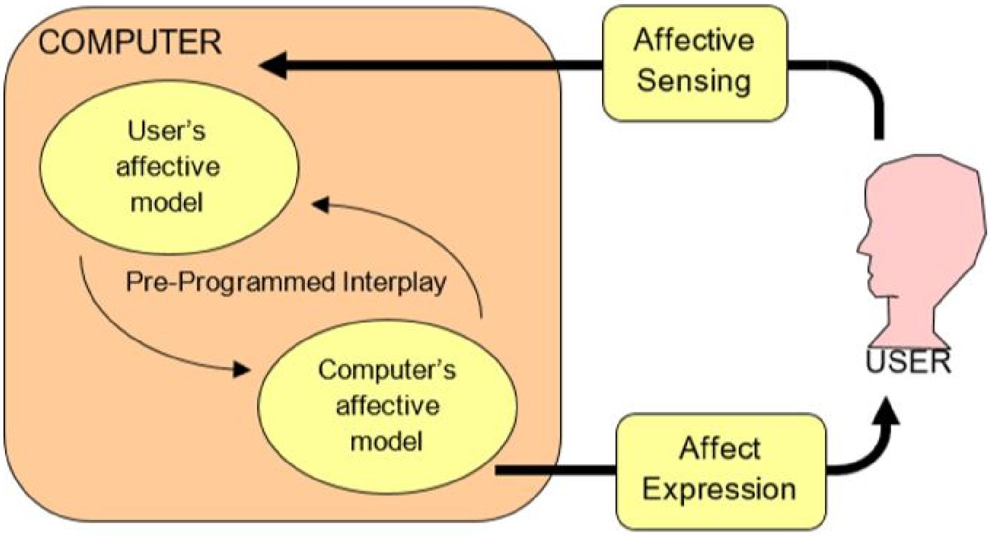


Figure 1.1: Simplified diagram showing the interaction between the key processes in affective computing identified in [Hud03]. (Diagram reproduced from [Bar08]

of the process, however, may involve some of the major challenges that must be overcome for the implementation of a fully-functional affective computing system. In fact, Picard identified Sensing and recognizing emotion as one the key challenges that must be conquered to bring the full promise of affective computing concepts to fruition [Pic03] and this topic is what this dissertation is focusing on to improve the emotional-perception capabilities of computers. Although the topic of affective computing was introduced two decades ago, the progress in this field is not as advanced as compared to other fields in artificial intelligence, due to many reasons, for example, the lack of interest or previous lack of the necessary real-time computational power. Nonetheless, recent developments in related fields, such as machine learning, big data, and computer vision have reached a level where it is possible to attempt the actual implementation of affective systems. Especially, correct machine learning advances have significantly re-defined how to automate computers to learn by themselves.

In the pursuit of solutions for that important challenge, there have been many approaches proposed. Specifically, a wide variety of mechanisms have been suggested for affective sensing. Some research groups have attempted the assessment of user affective states using streams of data that are commonly available in contemporary computing systems, such as video from the users face, audio from the users voice and text typed by the user on the keyboard. Zeng et al. [ZPRH09], provided an interesting survey of relevant systems that use video and/or audio, to estimate the users affective state. Most vision-driven approaches are based in the known changes that occur in the geometrical features (shapes of the eye, mouth, etc.) [CHFT06] or facial appearance features (wrinkles, bulges, etc.) [GD05] of the subject, according to different affective states. Cowie et al. associated acoustic elements to prototypical emotions [CDCT+01]. Some other groups explored the coordinated exploitation of audio-visual clues for affective sensing [CDCT+01]. Liu et al. focused on the utilization of text typed by the user for affective assessment [LLS03]. Approaches in this area of work include Keyword Spotting (e.g., [Ell]); Lexical Affinity (e.g., [AOC03]); Statistical Natural Language Processing (e.g., [GSH+00]); etc.

Other groups have attempted to identify the physiological modifications that are directly associated with the affective states and transitions in human beings, and have proposed methods for sensing those physiological changes in ways that are noninvasive and unobtrusive to a computer user. The reconfiguration experimented by a human subject as a reaction to psychological stimuli is controlled by the Autonomic Nervous System (ANS), which innervates many organs and structures all over the body. The ANS can promote a state of restoration in the organism, or, if necessary, cause it to leave such a state, favoring physiologic modifications that are useful in responding to the external demands. In our case, ***the AffectiveMonitor system***, which is the focus of this dissertation, is our attempt to achieve the goal of empowering computers to recognize the user's emotion analyzing his/her facial expression and pupil diameter changes.

1.3 Research Questions and Hypotheses

Question: Will the proposed method provide a useful assessment of the users affective state, enabling it to react appropriately to it?

Hypothesis: By estimating the level of arousal and valence of the computer user via pupil diameter and facial expression, the computer will be able to place the users affective state in the Circumplex Model of Affect. Based on that estimation, a machine learning model could be used to synthesize an appropriate affective response by the computer to decrease, if necessary, the users negative feelings.

1.4 Outlines

This dissertation starts by explaining the methodology as well as the necessary background to understand the chosen approach in Chapter 2. Then the process of data acquisition will be outlined and details on how the human-subject experiment will be provided in Chapter 3. Chapter 5 will describe the AffectiveMonitor system in depth, including the details of its software, all its integrated features, the modules for data acquisition. Chapter 6 and 7 explain the method followed to build the model to classify the affective state of the user. Subsequently, the results and performance of the model will be reviewed in Chapter 8. Lastly, Chapter 9 will suggest future work and possible alternative ways to utilize the data collected for this research.

CHAPTER 2

METHODOLOGY

This chapter outlines the fundamental topics required as background knowledge for the explanation of the approaches developed for the method we propose for intrusive affect recognition in a computer user. The chapter starts by explaining the model of emotion called the Circumplex Model of Affect, which can be described briefly as a two-dimensional plot of representing arousal as a vertical axis and valence as a horizontal axis. Accordingly, in this model of affect, the affective state can simply be represented as a location specified by two parameters: arousal and valence. The rest of the chapter introduces background information on the mechanisms that might be used to determine those two parameters that characterize erg affective state of a computer user. The description will also outline the challenges encountered in assessing the affective parameters, and how this research sought to circumvent those challenges.

2.1 Model of Emotion

Early on, a persons affective state was typically mapped to a discrete system with a limited set of basic emotions and each emotion was considered independent of one another. Nowlis [NN56] reported his investigation and concluded that he thought there are between six to twelve monopole factors, based on the observation that those core emotions such as anger, fear, sad, happy, and so on can be distinguished separately by people regardless of their ethnic, age, or sex. Neurophysiologists believed that affective states can be treated as if they are each in a different dimension because each affective state has its own unique neural pathways in the Central Nervous System (CNS). Although theories of basic emotions were dominant in psychiatric and neuroscience research, the theory itself is based on speculation rather than on empirical observations. For example, one might assume that the emotion of sadness has an uncorrelated relationship with the feeling of happiness. However, there are some situations in which this theory of discrete emotion cannot give a clear explanation of how one feeling separates from another. For instance, the feelings of worry and fear are somewhat different but somewhat similar, at the same time. Thus, this theory still requires more substantial evidence to support it.

Later, Russell developed a theory that explains the affective state in more empirical terms. Russell has proposed The Circumplex Model of Affect [Rus80] as a model of the affective state of a human. The model is based on the fundamental idea that each affective state arises from the product of the interaction between two independent neurophysiological systems: arousal and valence. One good thing about Russells work is that he used the statistical tool of factor analysis in various psychological assessments he performed when he was conducting his experiments, and the results showed consistent outcomes that strengthen the support of his hypothesis. Russell had studied that English words used to refer to different type of emotions can be placed in scales to rate the degree of pleasure-displeasure and degree-of-arousal that they convey. He also found out that these dimensions are bipolar and each affective state could be arranged on the circumference of a circle in a two-dimensional space; which he named The Circumplex Model of Affect. In his model the pleasure-displeasure dimension is placed as the horizontal axis, where its negative pole represents displeasure and its positive pole is regarded as pleasure, On the other hand, the degree-of-arousal dimension is positioned as the vertical axis, where its negative pole corresponds to low in arousal and the positive pole is assigned to high in arousal. The arrangement of the affective states in the circumplex model depends on the way an affective word, is projected in the circumplex model of affect. For example, The word Bored appears to be projected near the center of

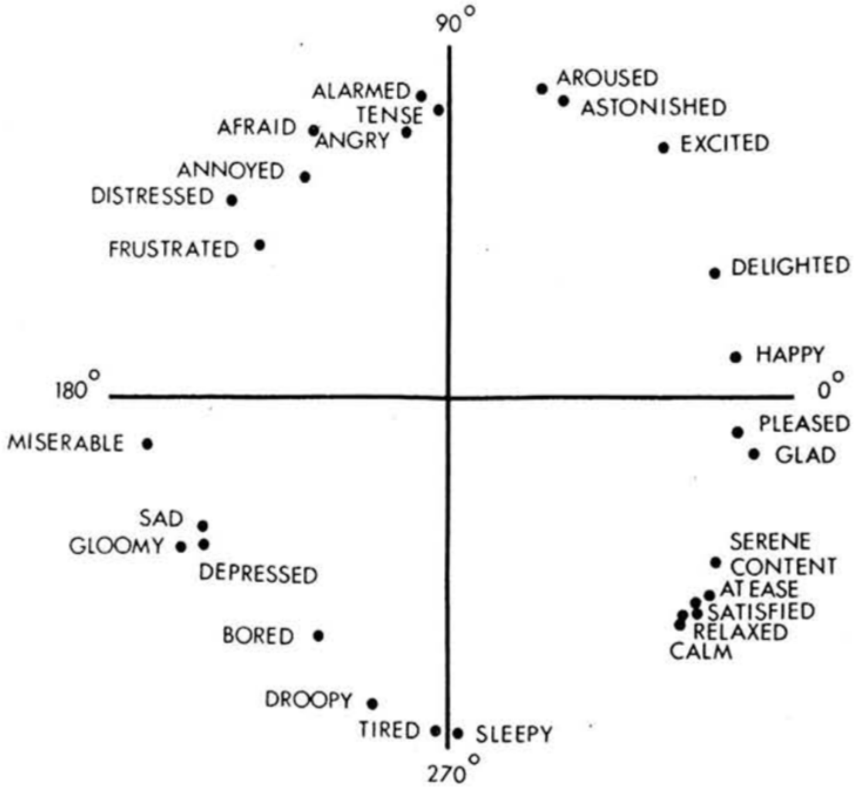


Figure 2.1: Direct circular scaling coordinates for 28 affect words (Figure from [Rus80])

the bipolar dimension of pleasure-displeasure (valence); and the same word appears to be placed in a region toward the negative pole of the degree-of-arousal dimension (arousal).

In his well-known study [Rus80] published in 1980, he proposed the placement of 28 affective words in the circumplex model of affect as shown in Figure 2.1. The first quadrant in includes locations with angles between 0°and 90°and the angles are considered to increase in the counterclockwise direction. Each quadrant can be briefly named according to the relationship between arousal and valence.

In this dissertation, we have chosen Russell's Circumplex model of affect as our hypothetical model, serving as the basis for affect characterization pursued in this study.

2.2 Assess the affective state

So far, we have explained the Circumplex Model of Affect where the location of each affective state is defined by two parameters: arousal and valence. The goal to assess the affective state of a computers user can be achieved if we can estimate these two parameters. One way to estimate the level of arousal and the level of valence from a computers user is to observe the changes of his/her bio-signal indicators which are directly affected by arousal and valence. Two indicators that we selected to observe are Pupil Diameter (PD) for arousal assessment and Facial Expression for valence assessment.

2.2.1 Arousal Assessment by Pupil Diameter

There are have been numerous studies in the neuroscience field that produced strong evidence for the identification of the segment of the nervous system which directly influences our reactions to psychological stimuli (e.g., arousal). This is the Autonomic Nervous System (ANS) (Figure 2.2).

The Autonomic Nervous System (ANS) coordinates the cardiovascular, respiratory, digestive, urinary and reproductive functions according to the interaction between a human being and his/her environment, without instructions or interference from the conscious mind [Riz15]. According to its structure and functionality, the ANS is studied as composed of two divisions: The Sympathetic Division and the Parasympathetic Division. The Parasympathetic Division stimulates visceral activity and promotes a state of rest and repose in the organism, conserving energy and fostering sedentary housekeeping activities, such as digestion [Riz15]. In contrast, the Sympathetic Division prepares the body for heightened levels of somatic activity that may be necessary to implement a reaction to stimuli that disrupt the rest

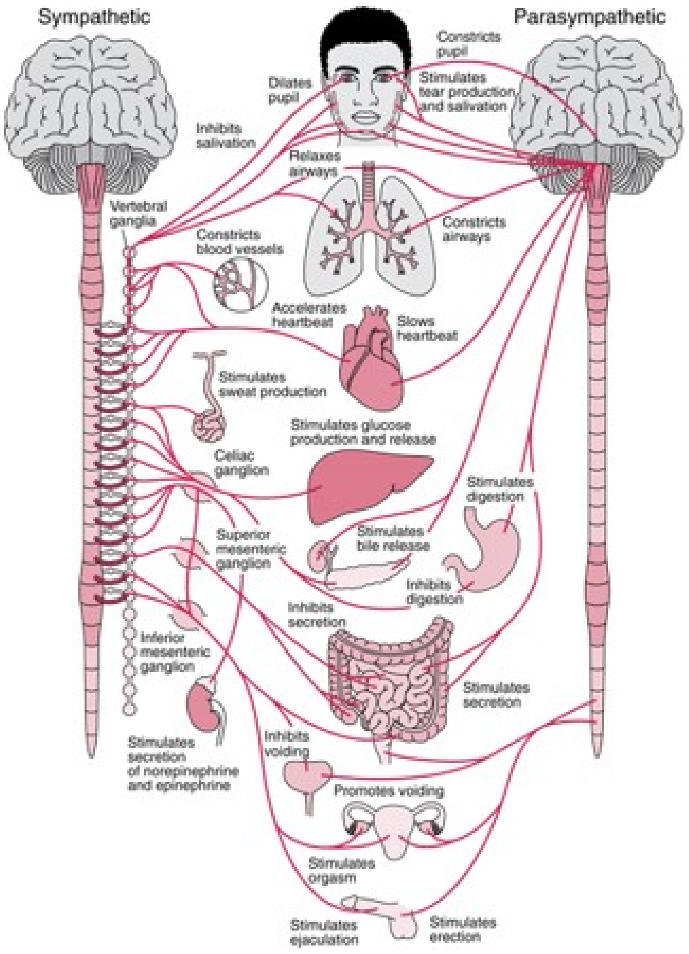


Figure 2.2: Autonomic Nervous System (ANS) (Picture from [Low])

and repose of the organism. When fully activated, this division produces a flight or fight response, which readies the body for a crisis that may require sudden, intense physical activity. An increase in sympathetic activity generally stimulates tissue metabolism, increases alertness, and, from a global point of view, helps the body transform into a new status, which will be better able to cope with a state of crisis. Parts of that re-design or transformation may become apparent to the subject and may be associated with measurable changes in physiological variables. Variations in sympathetic and parasympathetic activation produce physiological changes that can be monitored through corresponding variables, providing, in principle, a way to assess the affective shifts and states experienced by the subject. Parasympathetic and sympathetic activations have effects that involve numerous organs or subsystems, appearing with a subtle character in each of them. Therefore, one approach to affective sensing might be based on monitoring the changes in observable variables that are brought about by an imbalance in the sympathetic-parasympathetic equilibrium introduced by sympathetic activation. These changes can then be matched to the fundamental types of states for which each of these divisions of the Autonomic Nervous System prepares us (The sympathetic response prepares us for fight or flight, whereas the parasympathetic response sets us up for rest and response). Accordingly, the predominance of sympathetic activity can very well be taken as an indicator of arousal, represented on the vertical axis of Russells Circumplex Model of Affect [Rus80]. It is, indeed, common to experience acceleration of our heart rate (evidence of sympathetic activation) both, while we take a crucial test and when our favorite sports team is winning a match.

Much of previous work at the FIU DSP Laboratory has focused on signal processing methods to estimate a level of sympathetic activation using data recorded from non-invasive physiological sensors, such as Electro-Dermal Activity (EDA), also referred to as Galvanic Skin Response (GSR), and, most promising due to its complete unobtrusiveness, Pupil Diameter (PD) monitoring, using infrared video analysis (commonly used in eye gaze tracking, EGT equipment). Our approach to assessing the level of arousal experienced by the subject is through the monitoring of the pupil diameter, measured, in real time, by many eye gaze trackers (EGTs). This approach, in fact, targets the estimation of sympathetic activation (and simultaneous parasympathetic deactivation) in the Autonomic Nervous System (ANS). Previously, the FIU DSP Lab group has explored the monitoring of pupil diameter from a computer user, utilizing an ASL-504 eye-gaze tracker, which reports the estimated pupil diameter in pixels (integer values), for the assessment of affective states in the user [BZRG07]. This approach has a strong anatomical and physiological rationale. The diameter of this circular aperture is under the control of the ANS through two sets of muscles. The sympathetic ANS division, mediated by posterior hypothalamic nuclei, produces enlargement of the pupil by direct stimulation of the radial dilator muscles, which causes them to contract [SSCP04]. On the other hand, pupil size decrease is caused by excitation of the circular pupillary constriction muscles innervated by the parasympathetic fibers. The motor nucleus for these muscles is the Edinger-Westphal nucleus located in the midbrain. Sympathetic activation brings about pupillary dilation via two mechanisms:

(i) An active component arising from activation of radial pupillary dilator muscles along sympathetic fibers.

(ii) A passive component involving inhibition of the Edinger-Westphal nucleus [BrEu].

The rationale for arousal assessment on the basis of pupil diameter monitoring is also supported by other independent experiments in which pupil diameter has been found to increase in response to stressor stimuli. Partala and Surakka used sounds from the International Affective Digitized Sounds (IADS) collection [LBC99] to provide auditory affective stimulation to 30 subjects, and found that the pupil size variations corresponded to affectively charged sounds [PS03]. In our previous work from the FIU DSP lab group [GBA09a], it was verified that an enlargement of the pupil diameter is observed when the subject experiences sympathetic activation from exposure to stressor stimuli (incongruent Stroop word presentations), therefore providing further support for the rationale of the combined system described in this dissertation. Figure 2.3 shows some of the results obtained. In Figure 2.3,

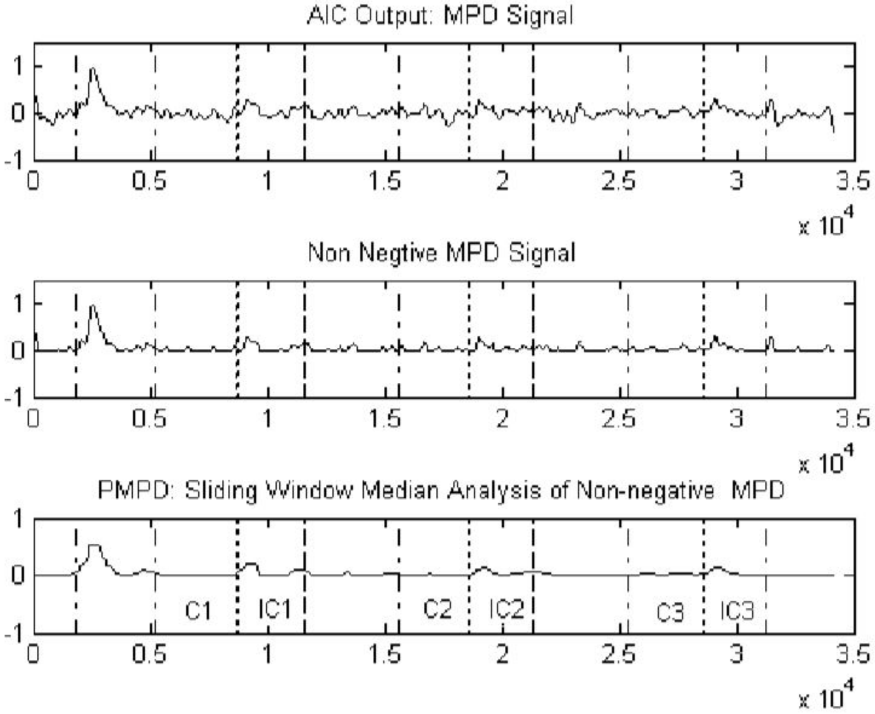


Figure 2.3: (From [GBA09a]) The bottom panel shows the increased in the Processed Modified Pupil Diameter (PMPD) signal, which correspond to application of stressor (Incongruent Stroop) stimuli, IC1, IC2 and IC3.

the elevations in the processed signal (PMPD), other than the initial transient at the beginning of the record, are seen to correspond with the intervals labeled IC1, IC2 and IC3, which were the intervals of the experiment when the subject was presented with incongruent Stroop word presentations. In conclusion, the pupil becomes dilated when a person experiences sympathetic activation (stress, aroused) while conversely, the pupil is constricted when his /her affective state is dominated by parasympathetic activation (peaceful, calm).

2.2.2 Valence Assessment by Facial Expression

In term of valence, psychologists define it as "any relatively brief conscious experiences characterized by intense mental activity and a high degree of pleasure or displeasure" (Quote from [Mat01b]). the elevations in the processed signal (PMPD), other than the initial transient at the beginning of the record, are seen to correspond with the intervals labeled IC1, IC2 and IC3, which were the intervals of the experiment when the subject was presented with incongruent Stroop word presentations. In conclusion, the pupil becomes dilated when a person experiences sympathetic activation (stress, aroused) while conversely, the pupil is constricted when his /her affective state is dominated by parasympathetic activation (peaceful, calm). [Dam05]. By observing transitions in the activity of organs of the human body, such as facial muscles, which occur as a result of emotional stimuli, we can classify human expressions of emotion or, in this case, identify the valence of those emotions. It has been proposed that the most basic and distinctive signs of experiencing emotions are the corresponding changes in facial expression. Even before we attempt to identify a person affective state from what he/she says, we instinctively observe another persons facial expression to determine what will be the appropriate interaction toward that person. In other words, we use our eyes to observe the changes in facial muscles that define facial expressions and then we interpret that expression based on the patterns we have seen in previous instances. Ekman noticed this fact and implemented the Facial Action Coding System (FACS) [EFA80], which provides a strong foundation for later studies in the affective computing field.

The Facial Action Coding System deconstructs the anatomic components of a facial expression into the specific Action Units (AU), and, accordingly, makes it possible to code the facial expressions of known affective significance on the basis of the contraction and relaxation of facial muscles. These associations can be leveraged in recognizing affective states from facial gestures. Humans do this through their intrinsic visual perception. For example, we may infer that a person is happy by observing the way the corners of his/her mouth are lifted, or the shape of his/her

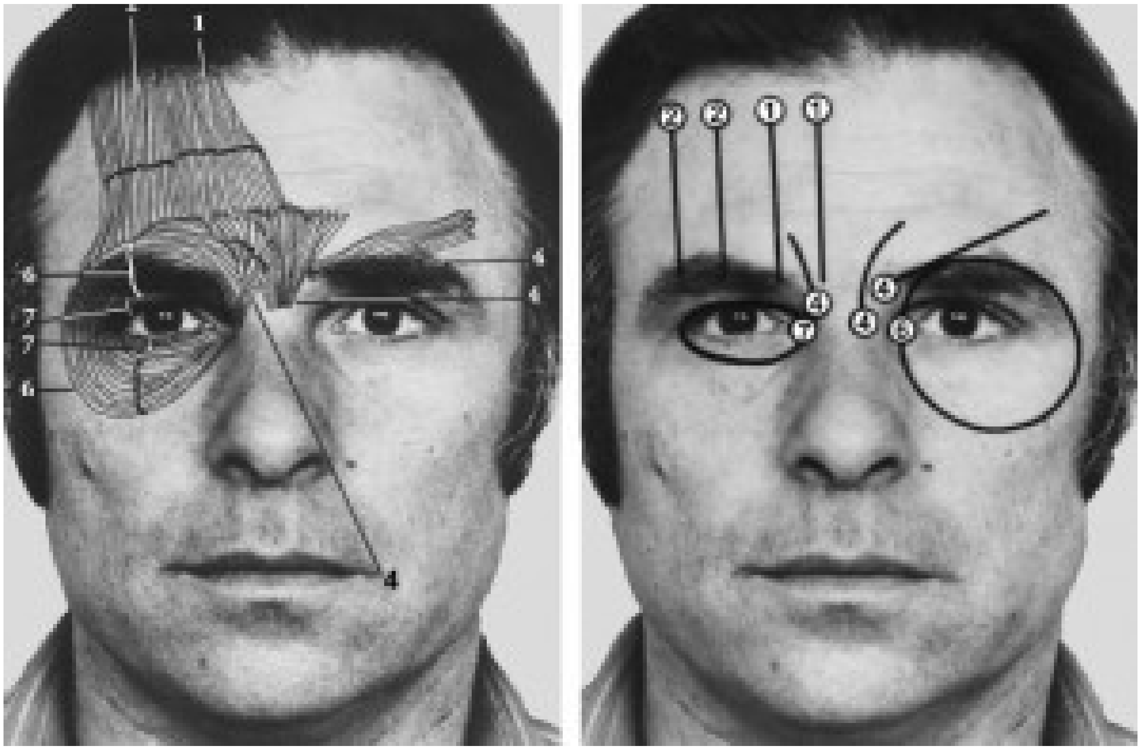


Figure 2.4: Examples of Action Units (AU) from Facial Action Coding System (FACS) (Picture from [EFA80])

eyes becomes narrower when a person smiles.

FACS provides a systematic way to encrypt the facial expression in an objective and compact set of standard parameters. By monitoring how the behavior of Action Units changes, corresponding to the different facial expressions, we can extract some unique patterns that can be used to classify types of facial expression. Ekman also suggests that the combination of the units can accurately make an inference about which kind of emotion the face is reflecting. Additionally, there have been extensive studies, conducted through decades, which reinforce this idea. Matsumoto et al. [Mat01b] have compiled a listing of the AUs typically activated in expressing 8 basic emotions (Table 2.1). In Figure 2.4, shows an example of the muscle activation that are used to define each Action Unit (AU). For example, AU4 (a.k.a. Brow Lowerer) represents the movement of facial muscles (Depressor Glabellae, Depressor

Supercilli, Corrugator), where the muscles constrict to attachment points (the places where the circular labels with the number 4 are. Notice that there are three muscles involved, hence there are 3 circular labels with the number 4 in them). The line extended from the circle refers to the placement of facial muscles associated with this AU.

Table 2.1: Action Units typically activated for 8 emotions. The numbers appearing in the table are referring to the index of Action Unit (AU) (Table is modified from [Mat01b])

Emotion	**AUs from Darwin's work**	**AU's from other human experiments**
Anger	4; 5; 24; 38	4; 5 or 7; 22; 23; 24
Contempt	9;10;22;41;61 or 62	12 ; 14
Disgust	10; 16; 22; 25 or 26	9 or 10; 25 or 26
Fear	1; 2; 5; 20	1; 2; 4; 5; 20; 25 or 26
Happiness	6; 12	6; 12
Joy	6; 7; 12	6; 12
Sadness	1; 15	1; 4; 15; 17
Surprise	1; 2; 5; 25 or 26	1; 2; 5; 25 or 26

Figure 2.5 summarizes the approaches that are followed in this research to characterize the affective state of a computer user. The research described in this dissertation pursues the assessment of the affective state of a computer user from two types of measurements. It will seek to estimate his/her arousal level and valence level by observing his/her pupillary response, influenced by the ANS, and by monitoring that persons facial expression, respectively. The following chapters, will explain in detail how this strategy was implemented practically.

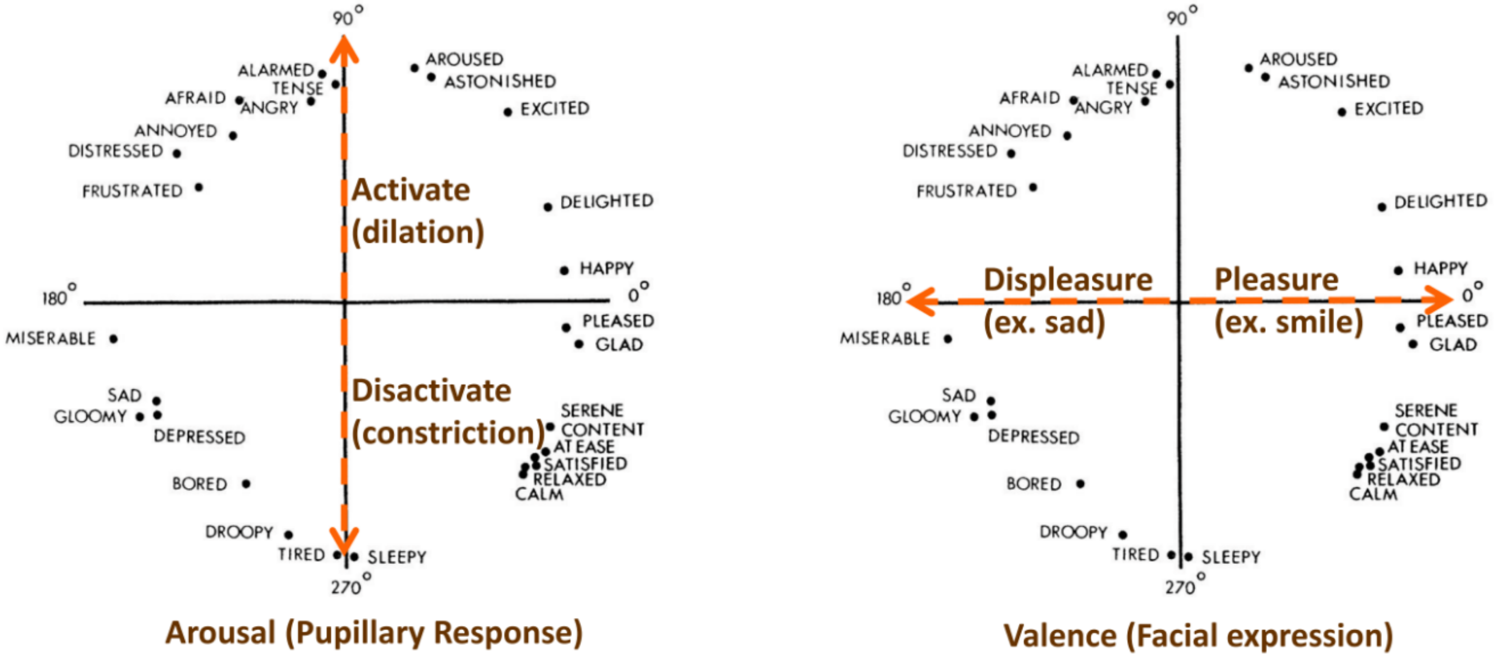


Figure 2.5: Summary of our methodology

CHAPTER 3

DATA ACQUISITION

The goal of this research is to build a supervised machine learning model to classify the computer users state of affect. One of the initial steps towards that goal is the identification of the types of data required for this task. The target data has to be able to reflect the changes of arousal and valence level of the user but, in addition, it is necessary that the data acquisition process should not interfere with the interaction between the user and the computer. Based on the Circumplex Model of Affect, there are two parameters we have to estimate to assess a users affective state: arousal and valence.

It is known that the pupillary response is influenced by the Autonomic Nervous System (ANS). The pupil is dilated (larger pupil diameter) when the user is in a high arousal state. Conversely, the diameter is constricted (smaller pupil diameter) when the user is in a low arousal state. Therefore, we can estimate the arousal level through the changes in the pupil diameter. Pupil diameter monitoring is also a good choice in terms of its non-intrusiveness during data acquisition.

For the assessment of valence case, the facial expression has long been considered a primary way for a human to observe another humans emotional changes, as well as a fundamental way in which humans express their emotion. We tend to conclude if the person we observe is happy if he or she is smiling and we can see that the person is sad if he or she is crying. Emotion is defined as the complex actions of a group of organs that are influenced by the mental activities and an associated high degree of pleasure/displeasure [Mat01a]. In our case, the group of organs that we are monitoring for affective valence assessment is the facial muscles. Therefore, pleasure/displeasure, i.e., affective valence, can be approximately estimated by monitoring of the subjects facial expression. The Facial Action Coding System (FACS) provides an empirical and systematic method to define the changes of facial muscles by detecting which Action Units (AUs) are activated during the changes in facial expression. In this research, the detection of facial changes is derived from changes in the 3D coordinates of the surface of the face. The data acquisition can be performed in a non-intrusive way by using the capabilities of the Kinect sensor module developed by Microsoft [Rah17].

Having identified the user variables to be monitored (pupil diameter and facial expression changes) the design of the data collection process plays an extremely important role to obtain appropriate data for the development of the affect recognition system. Thus, an experiment has been set up where human subjects will be presented with images from the International Affective Picture System (IAPS)[LP05] to elicit from them affective reactions, manifested through their involuntary changes in pupil diameter and in their facial expressions. In addition, subjects while also report the subjective assessment of their reactions through the Self-Assessment Manikin (SAM)[BL94].

During the recording sessions, a Kinect sensor was used to collect the 3D facial coordinates and the Facial Animation Parameter Units (FAPUs)[AA01] from the subjects face, as well as an estimate of the illumination level in the area around the eyes of the subject. Simultaneously, an Eye Gaze Tracking (EGT) system was used to record the pupil diameter in the eyes of the subject. The self-reports of arousal and valence marked by the subject in SAM for each IAPS image were also recorded into the dataset for later use.

The next section provides an explanation of the International Affective Picture System used as the stimulus for elicitation of affective responses in the subjects.

3.1 The International Affective Picture System

The International Affective Picture System (IAPS) is a large set of color photographs that elicit shifts in the subjects arousal and valence. IAPS contains a wide variety of stimulus types for more than 1,000 exemplars of human experience such as joyful, sad, fearful, attractive, angry, simple objects, scenery, etc. The idea is to present the subject with visual stimuli to modify his/her affective state while recording his/her reaction. The IAPS has been used across various fields of study to investigate emotion and attention worldwide and it is well-known for its replication and robustness. Pictures from IAPS are rated with arousal, pleasure, and dominance mean values, based on reactions from men and women, which make them suitable to be used as stimuli in this study. More in-depth information about IAPS can be found in [LP05].

For this research, IAPS provides both the stimuli (pictures) and the labels for the levels of arousal and valence needed for the design of a classifier under the supervised machine learning paradigm. IAPS provides us the mean and the standard deviation of arousal, valence, and dominance values, according to the ratings that thousands of subjects gave to the pictures in previous characterization studies. This means that the number of mean arousal and mean valence that comes with each picture has already reduced the potential bias of the rating by an individual, which may react differently based on their personal background, religion, culture, and etc. The mean values represent how the majority of people react to each particular picture. The IAPS documentation also provides the means and the standard deviations calculated for separate genders which could be very useful in case we include the gender as one of our features for training a predictive model.

The group that has developed the IAPS database suggests that the IAPS pictures should not be released to the public or be seen by the participants before the experiment starts to preserve their usefulness as emotional stimuli.

We chose to only use the arousal and valence mean values form the IAPS database, as the dominance parameter was not central to our approach, based on the Circumplex Model of Affect [Rus80]. In figure 3.4a, each point shows the location of the mean of arousal (Y-axis) and valence (X-axis) of each of the picture samples in the IAPS database. Each picture is represented by a circle. Additionally, the radius of each circle represents the standard deviation of the arousal rating across all participants. The tool used by the subjects who participated in the experiments performed to develop the IAPS database for rating these two parameters: arousal and valence, is called the Self-Assessment Manikin (SAM). This tool is further described in the next section.

3.2 Self-Assessment Manikin

The Self-Assessment Manikin (SAM)[BL94] is a tool for a non-verbal, pictorial assessment reporting technique that directly expresses the pleasure, arousal, and dominance associated with the affective state of the subject while being exposed to a stimulus. We mainly focus on the 2-dimensional Circumplex Model of Affect; therefore, dominance reactions are not considered. As demonstrated in figure 3.1, the SAM figure varies along each scale. In the arousal scale, the left-most figure corresponds to the most extremely stimulated, excited, frenzied, jittery, wide-awake, or aroused state. While the other end of the scale represents a completely relaxed, calm, sluggish, dull, sleepy, or unaroused state. The scale ranges from 1 to 9 for the purpose of intermediate fine-grained rating. For the pleasure (valence) assessment, the scale works the same way as for arousal except, in this case, the left-most figure represents a highly happy, pleased, satisfied, contented, hopeful state; while the opposite end represents a very unhappy, annoyed, unsatisfied, melancholic, despaired, bored state.

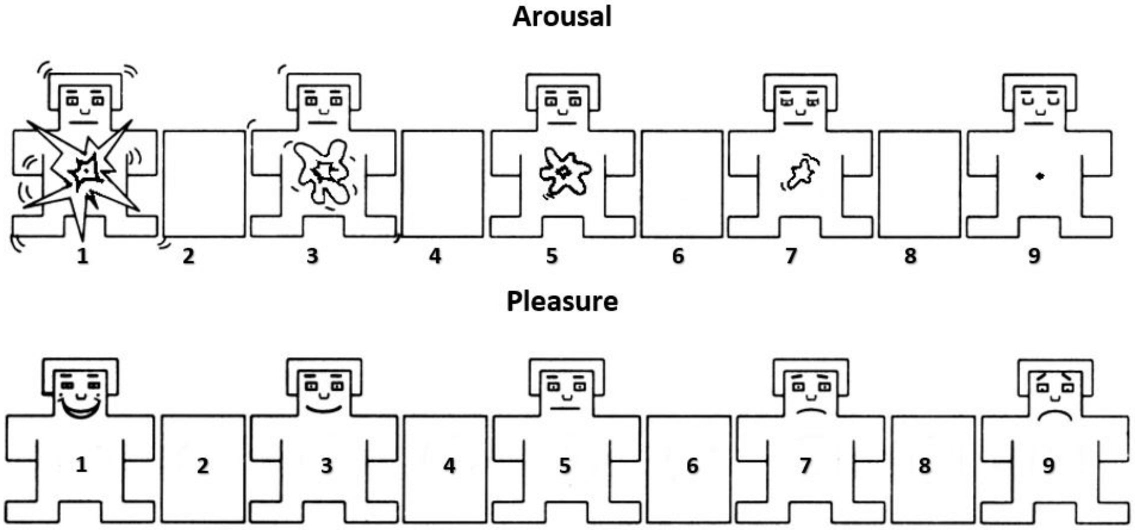


Figure 3.1: Self-Assessment Manikin (SAM) (from[LP05])

3.3 Experiment Setup

The entire data collection process is depicted in the diagram shown in figure 3.2. This diagram describes the process handled by the AffectiveMonitor application [TOlR$^+$18] and indicates the list of output files obtained from the data collection process. Kinect, running on the primary computer is responsible for obtaining 3D facial coordinates while the TM3 Eye-Gaze Tracker device running on a secondary desktop computer records the pupil diameter signals and sends them over to the primary machine. These data are recorded during the experiment session and are written out in a timely manner, for each frame, to output files. We show how the experiment has been set up and its environment in figure 4.1.

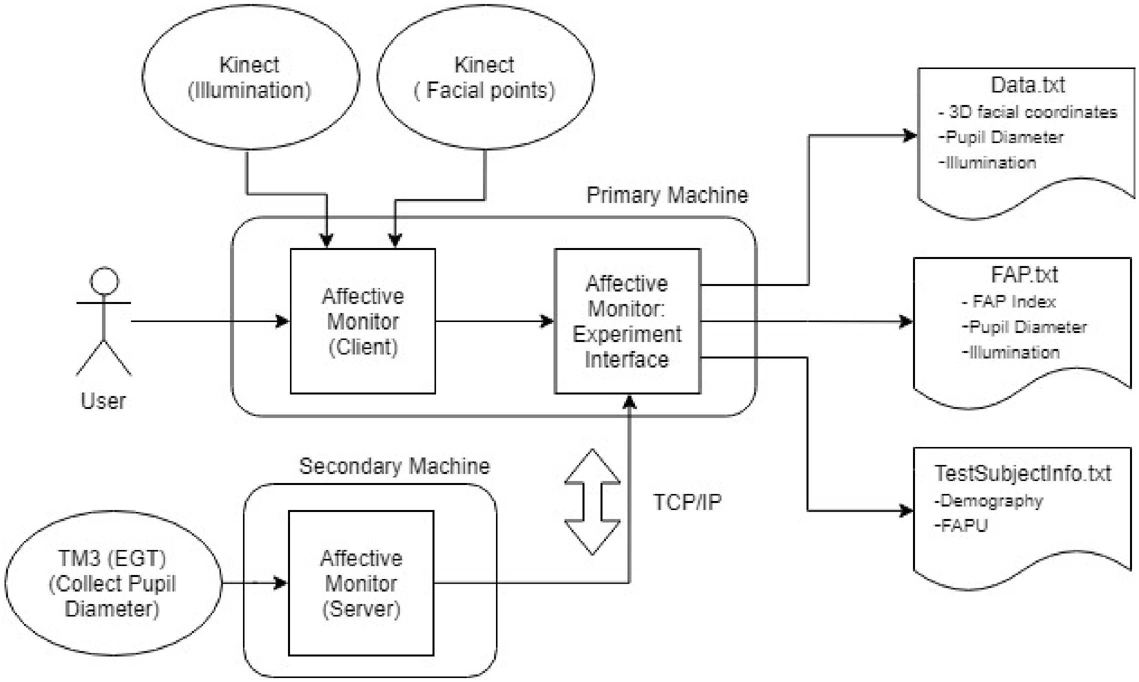


Figure 3.2: Bird's eye view of the system (Data collection process)

3.3.1 Experiment Procedure

AffectiveMonitor has a separate ”Experiment” interface tab section (Figure 3.3b) to conduct the experiment from beginning to end. The experiment takes about 35 minutes and before the experiment session begins, the subject will go through the following protocol for the purpose calibration.

1. Listen to the brief description of what the study is for and what the participant will be doing throughout the experiment.
2. Sign the form of consent to his / her participation in this study.
3. Go through the scanning process to adjust (customize) the shape of a 3D facial model.
4. The experimenter adjusts the position of the subject for adequate pupil diameter recording.

5. The subject will provide general information, such as his /her gender, age, and ethnicity, which is kept confidential and is not associated with the identity of the subject.

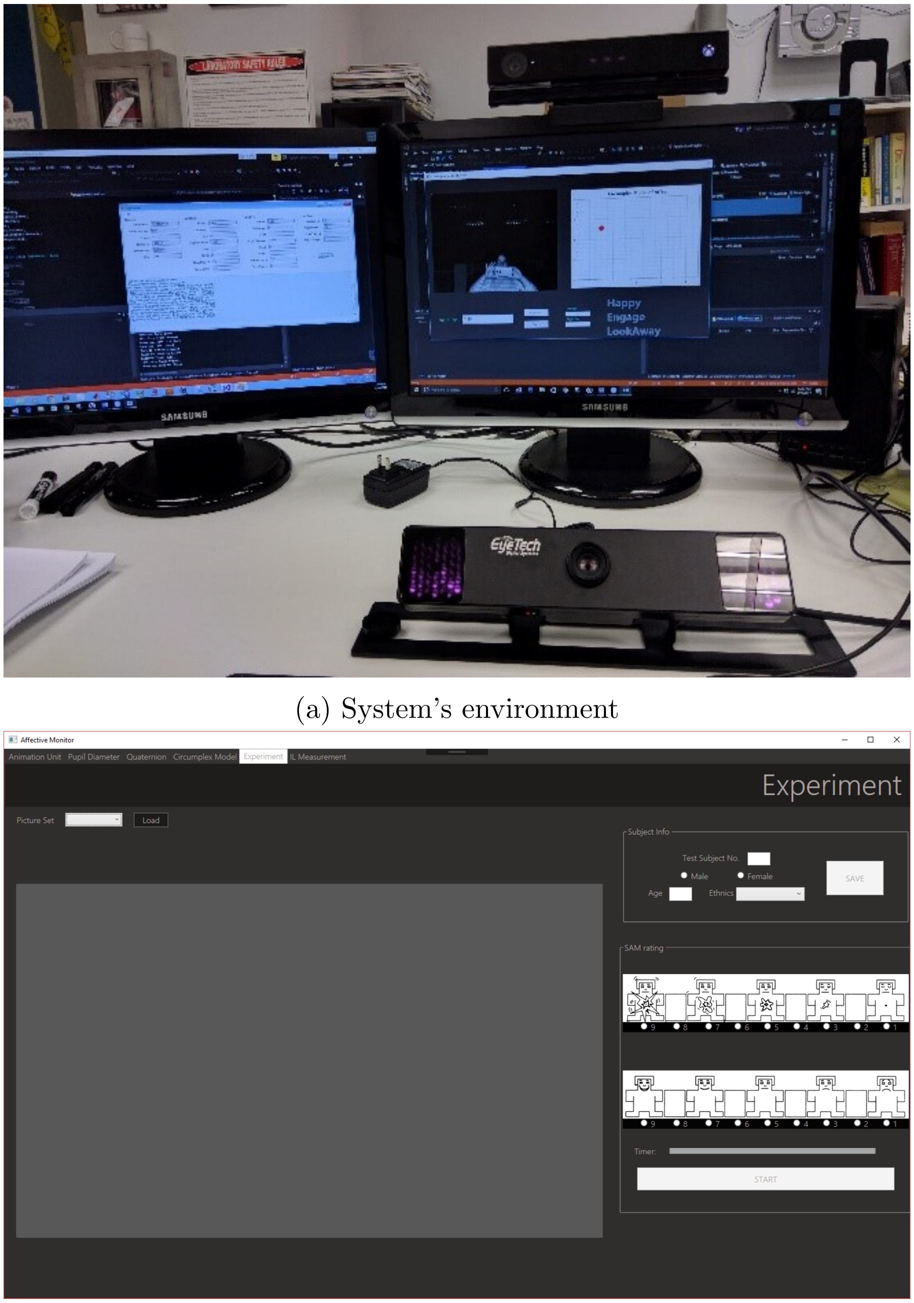
6. The data recording process starts.

During the experiment, 70 pictures selected from IAPS will be shown to the subject, one after another, until all samples are presented. For each sample, the subject is asked to look at the picture for 6 seconds, then immediately after, rate their affective state assessment via SAM (5 seconds). In between samples, a gray screen is shown during the resting period. The subject is urged to stay still during the first 6 seconds when he/she is first presented with the stimulus in order to reduce the measurement interference that could occur during the recording process.

The experiment is conducted in a relaxed environment where the participant can focus on looking at the pictures and providing the corresponding ratings of arousal and valence, using the SAM tool.

3.3.2 Sample Selection

For the experiment, we selected IAPS pictures on the basis of the mean and variance of arousal and valence that come with each picture from the IAPS repository. Our criterion for selecting the samples is based on the study of a 12-Point Affect Circumplex (12-PAC) model of Core Affect [YRS11] which is also based on the Circumplex Model of Affect. This study refined Russells framework by hypothetically dividing the Circumplex model into twelve segments called the 12-Point Affect Circumplex (12-PAC) structure. By finding the correlation between many previous studies and their own, the authors report their analysis and their placement of moods on a 12- PAC structure as shown in figure 3.4b. Based on this study, we selected the



(a) System's environment

(b) AffectiveMonitor: Experiment Interface

Figure 3.3: An entire system including Kinect V2 (on top of the screen) and TM3 (in front of the computer) is shown in Fig. 4.1. Fig. 3.3b shows an experiment interface of the AffectiveMonitor application

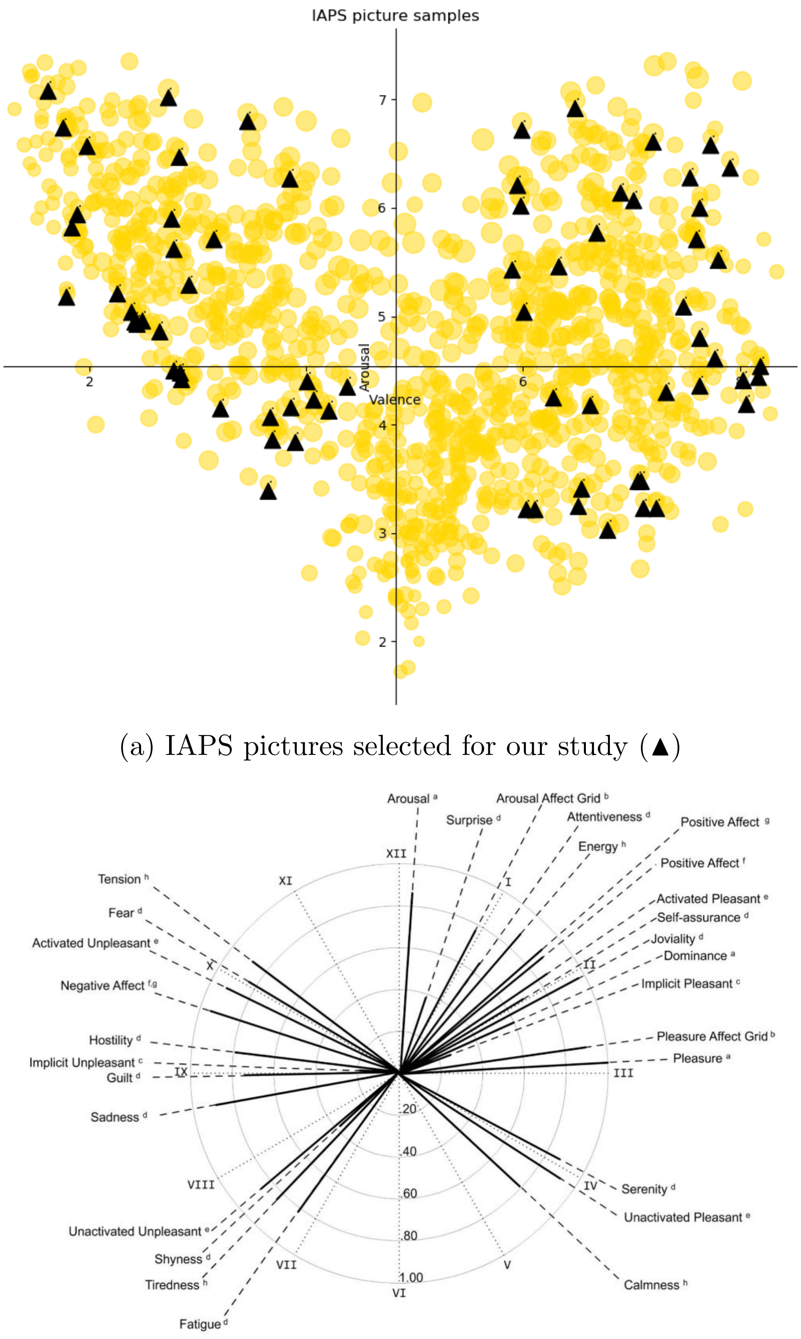
IAPS samples that are located around desired angles of those core affects that have more than 60% likelihood to appear in the Circumplex Model on that angle. That behavior is characterized by the length of the solid line depicted in figure 3.4b for each of the mood scales, based on studies conducted to quantify the level of arousal and valence corresponding to those individual moods. Even though there are 28 moods in total plotted around a circle in Russells model, we are primarily focusing on classifying the users affective state at a coarse scale first, before we move on to a more fine-grained classification. So, our initial goal is to build the predictive model that can classify the affective state into roughly 4 classes. These 4 classes could then be related to the 4 quadrants in the Circumplex Model of Affect and interpreted correspondingly:

1. Positive Arousal and Positive Valence (Quadrant 1)
2. Positive Arousal and Negative Valence (Quadrant 2)
3. Negative Arousal and Negative Valence (Quadrant 3)
4. Negative Arousal and Positive Valence (Quadrant 4)

Accordingly, we selected 70 IAPS samples, including pictures from each of the four quadrants, which are associated with 10 types of mood, as shown in table 3.1. The selection is structured in a way that balances the distribution across the circumplex model so we can build the predictive model with balanced data.

3.4 Data Acquisition

In this section, we explain the method of obtaining the measurements that are used to generate the features used for affect classification, including, 3D facial coordinates, pupil diameter, Facial Animation Parameters (FAPs), and illumination around the



(a) IAPS pictures selected for our study (▲)

(b) Core Affects on 12-PAC structure (from[YRS11])

Figure 3.4: Fig. 3.4a shows a plot of means of arousal and valence for images in the IAPS repository on top of the Circumplex Model of Affect. Notice that the radius of each plotted circle varies according to its variance. The triangular labels indicate the images chosen for use samples in this experiment. Fig. 3.4b demonstrates thirty mood scales which are placed within the 12-PAC structure with CIRCUM-extension method[YRS11]. The length of the solid line from the center can be roughly described as the maximum likelihood of placing a mood on the designated angle

.

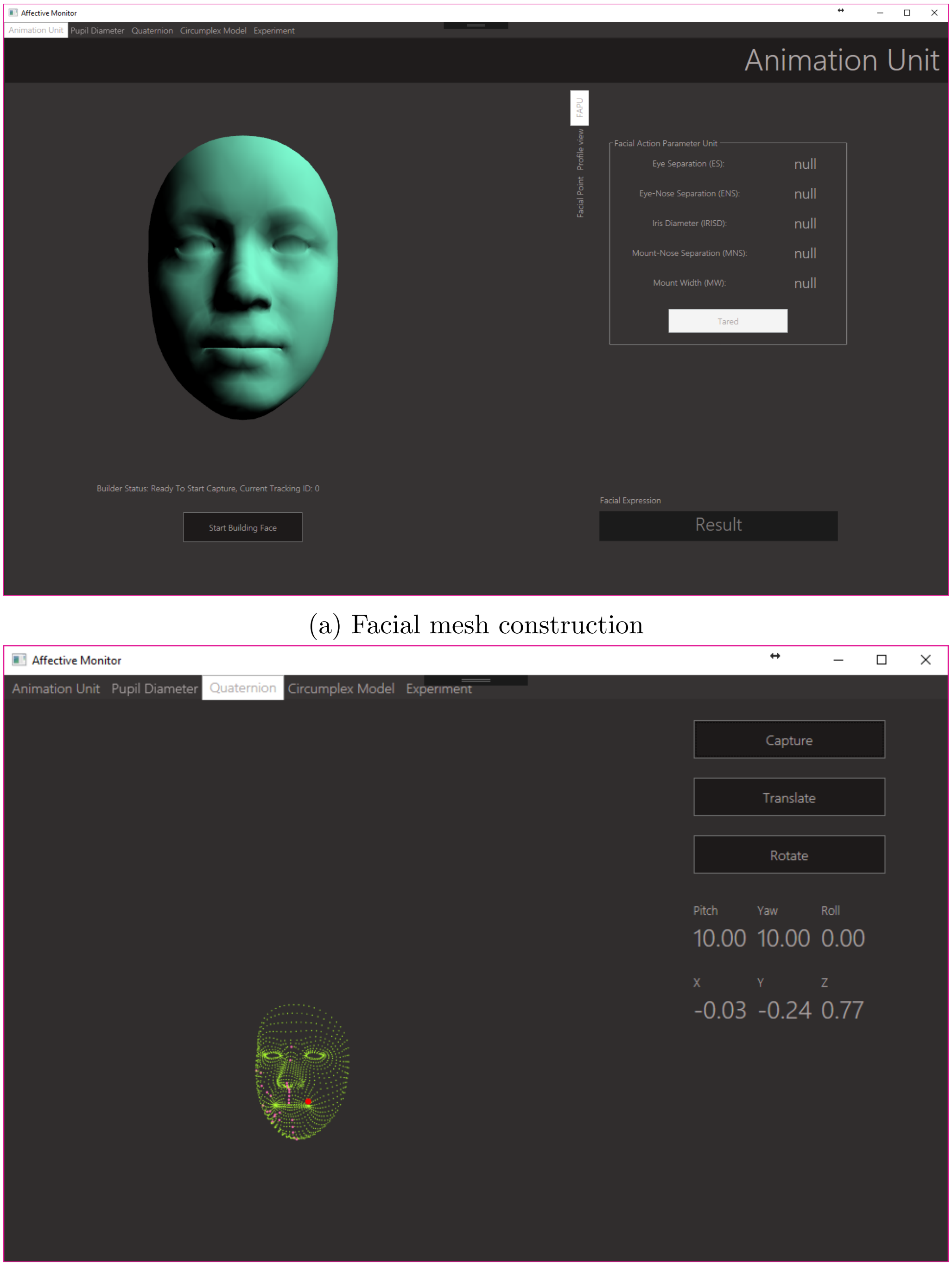
Table 3.1: Selected Samples listed by Picture ID from IAPS

	Pleasure	Joviality	Attentiveness	Disgust	Fear	Negative	Sadness	Tiredness	Calmness
1	1440	8499	4664	9301	9252	9007	2456	2399	5811
2	2550	8501	4604	7359	9413	9320	2095	2039	5870
3	2260	8080	4689		9940	9342	2301	2752	1604
4	2070	7600	8179		6550	9295	2141	9390	5875
5	5831	7451	8490		2981		2799	9913	1419
6	7200	2092	4574		9491		4598	9395	2000
7	2154	8300	4232		9042			9190	5410
8	2151	8200	5950		6250			2400	7325
9	5910	4626	1050		9325			2695	5725
10		8540	5972		9433			4635	

eyes of the subject. All of them are recorded with the same timestamp by the AffectiveMonitor application.

3.4.1 3D Facial Coordinates

Kinect has provided the basic software framework, called HD face [Rah17]], that is needed to capture 3D coordinates of the surface of the face of the subject. This framework can detect the face of the closest person in front of the Kinect sensor and generate the persons 3D facial mesh model in real-time. Another interesting prospect of this framework is its ability to reconstruct the persons face shape by 3D scanning. We have integrated this framework into our AffectiveMonitor application to benefit from all the functionality that Kinect has to offer. The mesh model can also be represented by 3D coordinates and can be thought of as markers attached on the subjects face so whenever the subjects facial expression changes, the markers also move according to the corresponding facial muscle movement. By recording frame by frame, we can observe the changes in 3D facial coordinates that occur because of the subjects facial expression.



Eye Separation (ES): null

Eye-Nose Separation (ENS): null

Iris Diameter (IRISD): null

Mount-Nose Separation (MNS): null

Mount Width (MW): null

Builder Status: Ready To Start Capture, Current Tracking ID: 0

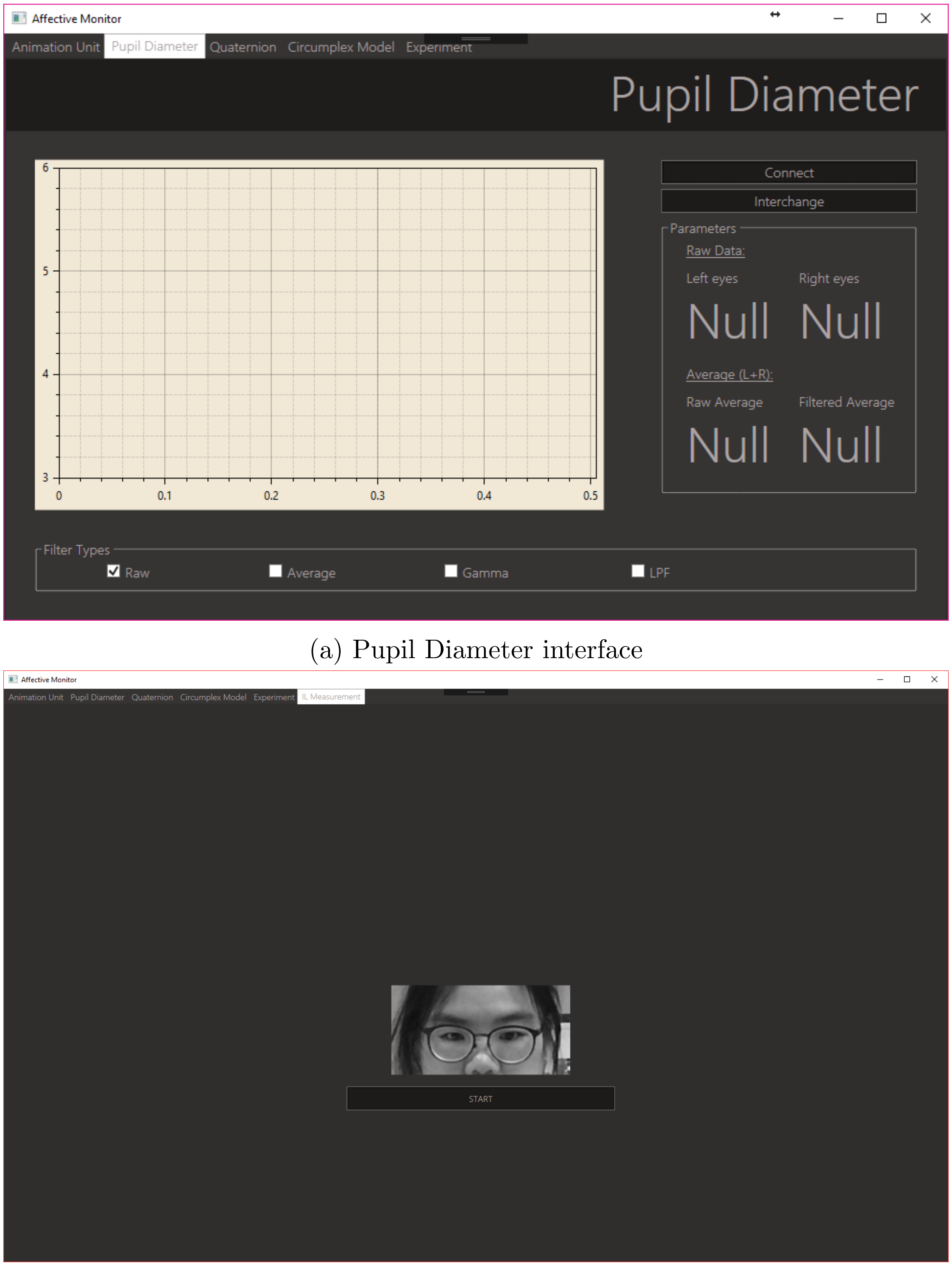
(a) Facial mesh construction

Pitch	Yaw	Roll
10.00	10.00	0.00

X	Y	Z
-0.03	-0.24	0.77

(b) Re-positioning and re-orienting facial points

Figure 3.5: Fig. 3.5a shows the interface of AffectiveMonitor for mesh construction. Fig. 5.5 displays the interface of AffectiveMonitor used for resetting the facial point cloud to its neutral position. The interface shows the shift in position and orientation in the Euclidean domain.



(a) Pupil Diameter interface

(b) Cropped video

Figure 3.6: Fig. 3.6a shows the interface of AffectiveMonitor for dynamic plotting of pupil diameter. Fig. 3.6b shows the cropped video used for illumination measurement around the eyes.

One problem that arises during the design of the experiment is the impossibility to completely restrain the movement of the subjects during the experiment. Body shifts can alter the position and orientation of the subjects face, which may complicate their processing. To circumvent this issue, we have built a feature in AffectiveMonitor to artificially re-position and re-orient the subjects face before recording the values.

Fortunately, Kinect also provides the pivot point, which is the centroid of the facial model, as well as the orientation (in quaternion format) of the face. Thus, we can reverse the rotation and transform the point cloud to a neutral position, a reference to the origin of the coordinate system by applying a coordinate transformation to each frame captured, on the basis of the available orientation of the face, to revert the rotation that may have occurred during the recording.

The Quaternion Inverse of a Rotation

A quaternion is a hyper-complex number of rank 4 that fulfills certain rules. Quaternions were introduced by Hamilton in 1843. They are widely used to represent orientations and rotations of three-dimensional objects to avoid the problem of gimbal lock [nas]. Equation 3.1 shows the mathematical notation for a quaternion, where q_0 is the scalar part of the quaternion while $\boldsymbol{q}$ is the vector part of the quaternion. A quaternion can also be represented by its components (q_0, q_1, q_2, q_3), where $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ are the standard orthonormal basis of $\boldsymbol{R}^3$, as a 4-tuple of the real numbers, as demonstrated in equation shown in Equation 3.2.

$$q = q_0 + \boldsymbol{q} = q_0 + \boldsymbol{i}q_1 + \boldsymbol{j}q_2 + \boldsymbol{k}q_3 \tag{3.1}$$

$$q = (q_0, q_1, q_2, q_3) \tag{3.2}$$

In this study, we are interested in using quaternion manipulations as a rotational operator that has properties that are well suited for this type of application. Below there is a brief description of how to apply the quaternion rotation operator to our 3D facial points collected with the Kinect module. For a more detailed explanation, please refer to [Kui99]. The quaternion rotation operator (L_q) is the result of the triple quaternion products that have a property to rotate the input vector ($\boldsymbol{v}$) to a resulting output vector ($\boldsymbol{w}$) around a quaternion axis in $\boldsymbol{R}^3$. From what we described, the operator can then be defined by the equation 3.3 where q^* is a conjugate of q.

$$\boldsymbol{w} = L_q(\boldsymbol{v}) = q\boldsymbol{v}q^* \tag{3.3}$$

A simpler computational formula for this process is indicated in Equation 3.4:

$$L_q(\boldsymbol{v}) = ({q_0}^2 - \boldsymbol{q}^2)\boldsymbol{v} + 2(\boldsymbol{q} \cdot \boldsymbol{v})\boldsymbol{q} + 2q_0(\boldsymbol{q} \times \boldsymbol{v}) \tag{3.4}$$

To have a better view on how to apply the quaternion rotation operator to an existing input vector $\boldsymbol{v}$, please observe Figure 3.7. In a very high-level explanation, one can think of the operation in equation 3.3 To have a better view on how to apply the quaternion rotation operator to an existing input vector $\boldsymbol{v}$ is rotated through an angle of 2θ about $\boldsymbol{q}$ as the axis of rotation. The angle (θ), in particular, is determined by the quaternion (q) itself. $\boldsymbol{a}$ is the component of $\boldsymbol{v}$ along the direction of q and $\boldsymbol{n}$ is the component of $\boldsymbol{v}$ along the direction of $\boldsymbol{v}$. In this case, $\boldsymbol{m}$ is the result of the quaternion rotation operation applying to $\boldsymbol{n}$ (that is, $L_q(\boldsymbol{n})$); and since $\boldsymbol{m}$ is the component of $\boldsymbol{w}$, we can say that $\boldsymbol{w}$ is the vector resulting from applying the quaternion rotation operation to $\boldsymbol{v}$. More detail explanation can be read in [Kui99].

To apply the appropriate rotation to our application, we first have to determine what elements in our problem represent each component in Equation 3.4. The

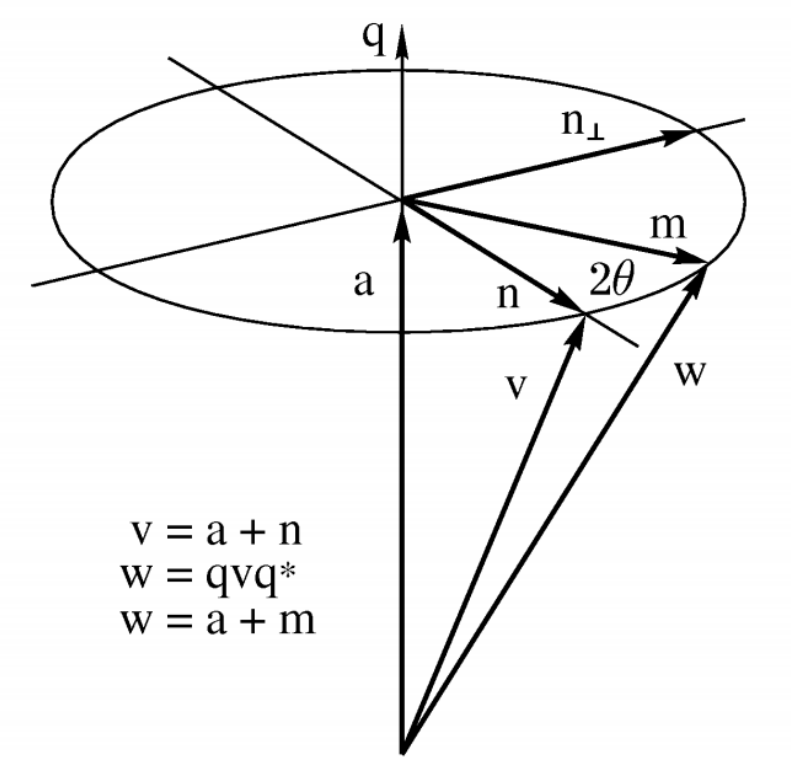


Figure 3.7: Quaternion Rotation Operator Geometry. (figure from [Kui99])

intent is to rotate the facial points model to the neutral position and to ensure that the captured face will be oriented as if the subject were looking directly to the Kinect module. Fortunately, the Kinect library framework provides the angle and the position of the face model in quaternion form. The information is attached to the pivot point of the face. We can place the face in a neutral position by moving the pivot point to the origin and adjusting its angle. The face model is composed of a group of 3-dimensional coordinates structured in the face shape with the common reference point (pivot point). Thus, we can consider one 3-dimensional coordinate as a vector input that will be rotated by the quaternion rotational operator, and then apply to the rest of 3-dimensional coordinates. To apply the quaternion rotation operator, we first have to find $\boldsymbol{v}$, which we can obtain by determining the vector from that particular point to the pivot point. For q, we already have it from Kinect at the pivot point; nevertheless, we want to invert the rotation back to its neutral position so we have to find the inverse of the quaternion provided by Kinect and then we can apply that inverse quaternion to the vector input. In our implementation of Equation 3.4. q_0 is the magnitude part of q and $\boldsymbol{q}$ is its vector part.

Another matter that we have to deal with is where the origin point of the quaternion vector, or in other words, the magnitude of a. If a is not starting from the origin then the rotated face model will also move in position; while our goal is to invert the angle of the face model to be in a neutral angle in place. (see Figure 3.7). That is why we have to transform the face models pivot point to the origin point first before we can apply the quaternion rotation operation. See the result in Figure 5.5

3.4.2 Pupil Diameter and Illumination

To acquire pupil diameter signals, we utilize the TM3 Eye-Gaze Tracker (EGT), which has the capability to measure the pupil diameter using the dark-pupil method. We set the sampling interval at 0.33s and average samples in an average window of 30-sample width. The pupil diameter signals are then transferred to the primary machine via TCP/IP, over ethernet cable. AffectMonitor has a feature to plot the average of the pupil diameter dynamically as shown in Figure 3.6a.

Many studies have shown that the pupil diameter is under the influence of the Autonomous Nervous System (ANS) and can be used as a marker for arousal level [GBA09b]. Unfortunately, pupil diameter is also susceptible to the amount of light impinging on the retina. To bypass this issue, we perform post-processing to address the effect of the pupillary light reflex on the pupil diameter values recorded. In order to account for this effect, the level of illumination around the eyes of ht subject must also be recorded as one of the output parameters. We obtain the illuminance measurement utilizing Kinects RGB camera by cropping the video around the eye area (Figure 3.6b) and calculating the illumination based on the cropped video. A more detailed explanation of the approach followed to address this challenge will be presented in Chapter 4.

Table 3.2: Facial Animation Parameter Unit (FAPU)

Description		FAPU Value
IRISD0 = 3.1.y 3.3.y = 3.2.y 3.4.y	Iris diameter (by definition it is equal to the distance between upper ad lower eyelid) in neutral face	IRISD = IRISD0 / 1024
ES0 = 3.5.x 3.6.x	Eye separation	ES = ES0 / 1024
ENS0 = 3.5.y 9.15.y	Eye - nose separation	ENS = ENS0 / 1024
MNS0 = 9.15.y 2.2.y	Mouth - nose separation	MNS = MNS0 / 1024
MW0 = 8.3.x 8.4.x	Mouth width	MW = MW0 / 1024

3.4.3 Facial Animation Parameter

The Facial Animation Parameter (FAP) is one concept of the components in MPEG-4 Face and Body Animation (FBA) International Standard (ISO/IEC 14496 -1 & -2) [PF03]. It describes a standard protocol to encode the virtual representation of human and humanoid movement, specifically around the facial region of the body. FAPs are commonly used to describe basic actions of facial expression for a synthetic face; for instance, in the CANDIDE model [AA01]. The ability of FAPs to encode the primitive expression information with small memory usage makes them interesting as an alternative method to record the subjects facial expression.

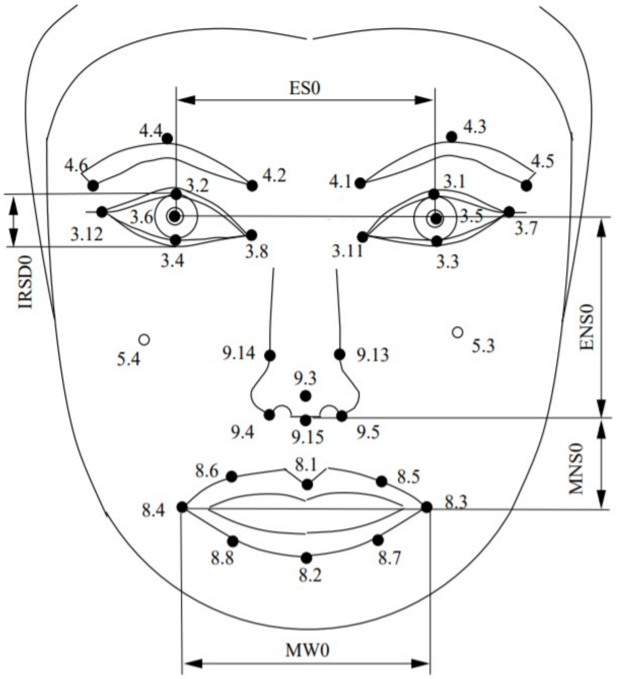


Figure 3.8: Facial feature points and Facial Animation Parameter (FAPU) (from[ZJZY08])

The Facial Animation Parameters (FAP) are defined by the displacement between facial feature points defined by the FBA standard (See Figure reffig:fapu) which are measured by Facial Animation Parameter Units (FAPUs).

FAPUs are normally calculated from a neutral face and divided by 1024 so that the unit is small enough to enable FAPs to be represented in integer numbers. The purpose of FAPUs is to allow a consistent way to interpret FAP indices for any facial model regardless of their shape and dimension. The description of the FAPUs and how to calculate them are listed in Table 3.2. We decide to output 19 FAPs listed in Table 3.3 which are actively related to basic facial expressions as desired output from the total of 68 FAPs [ZJZY08]. Note that in Figure. 3.8, the numbering of the facial feature points is according to FBA standard, while the index coordination system from Kinect is in a different listing. See Table 3.3 for the correspondence

Table 3.3: FAP Measurement with Facial Feature Points (FBA & Kinect)

FAP index	FAP Name	Distance of two feature points (FBA)	Distance of two feature points (Kinect)	FAPU
31	raise_l_i_eyebrow	Dy(4.2, 3.8)	Dy(346, 210)	ENS
32	raise_r_i_eyebrow	Dy(4.1, 3.11)	Dy(803, 843)	ENS
35	raise_l_o_eyebrow	Dy(4.6, 3.12)	Dy(140, 469)	ENS
36	raise_r_o_eyebrow	Dy(4.5, 3.7)	Dy(758, 1117)	ENS
37	squeeze_l_eyebrow	Dx(4.4, 3.8)	Dx(222, 210)	ES
38	squeeze_r_eyebrow	Dx(4.3, 3.11)	Dx(849, 843)	ES
19	close/open_t_l_eyelid	Dy(3.6, 3.2)	Dy(241, 1104)	IRSD
20	close/open_t_r_eyelid	Dy(3.5, 3.1)	Dy(731, 1090)	IRSD
41	lift_l_cheek	Dy(5.4, 3.12)	Dy(458, 469)	ENS
42	lift_r_cheek	Dy(5.3, 3.11)	Dy(674, 117)	ENS
61	stretch_l_nose	Dy(9.14, 3.8)	Dy(210, 1170)	ENS
62	stretch_r_nose	Dy(9.13, 3.11)	Dy(843, 1162)	ENS
59	raise/lower_l_cornerlip_o	Dy(8.4, 3.12)	Dy(91, 469)	MNS
60	raise/lower_r_cornerlip_o	Dy(8.3, 3.11)	Dy(687, 117)	MNS
53	stretch_l_cornerlip	Dx(8.4, 9.15)	Dx(91, 14)	MW
54	stretch_r_cornerlip	Dx(8.3, 9.15)	Dx(687 14)	MW
5	raise/lower_b_midlip	Dy(8.2, 9.15)	Dy(8, 14)	MNS
4	lower_t_midlip	Dy(8.1, 9.15)	Dy(19, 14)	MNS
3	open_jaw	Dy(8.2, 8.1)	Dy(19, 8)	MNS

between Kinect's index coordination system and FBA's coordination system.

3.5 Summary

Our goal is to collect the data suitable to train a supervised machine learning model, to classify the affective state of the subject in the Circumplex Model of Affect. In order to achieve that, we have to estimate two parameters: arousal and valence, with our model. In the case of arousal, we have found strong evidence supporting the notion that the pupil diameter is influenced by the Autonomous Nervous System, which is responsible for the state of arousal. In the case of valence, we decided to estimate this parameter on the basis of the subjects facial expression since pleasure and displeasure are directly expressed naturally by the activity of the facial muscles. Two data formats representing facial expression are recorded, 3D facial coordinates and Facial Animation Parameter indices and each has pros and cons. 3D coordinates are practical because they preserve the whole information recorded in the facial expression without losing any; while, FAPs are better with respect to memory usage. Other data that are collected along during the experiment, such as illuminance around the eye area, the distance between the subjects face and the Kinect sensor, and FAPUs, as they are necessary for scaling adjustment and calibration. Data are obtained in a time-stamped manner where pupil diameter, FAPs, 3D facial coordinates, and others are captured simultaneously and recorded together. Additionally, they are recorded in a customized output file for facilitating the transfer of the data to the analysis phase.

CHAPTER 4

REMOVE PUPILLARY LIGHT REFLEX

One effect that prevents the direct measurement of the Pupillary Affective Response (PAR) from the raw pupil diameter signals is the Pupillary Light Reflex (PLR). This chapter describes this effect and possible approaches to remove the effect of the Pupillary Light Reflex (PLR) component from pupil diameter signals obtained by an Eye-Gaze Tracking device (Eyetech Digital TM3) using the RGB camera from the Kinect module as a way to measure the illuminance around the eyes of the user. The purpose of this study is to obtain pupil diameter signals that mainly reflect the Pupillary Affective Response (PAR) used to estimate the arousal level in the response of a human subject to affective stimuli. One previously proposed approach includes using an Adaptive Interference Canceller (AIC) technique to filter out the Pupillary Light Reflex (PLR) from pupil diameter signals (PD). We also present the empirical method followed to replace a stand-alone light meter with the RGB camera from Kinect to measure illuminance.

4.1 Introduction

Previous research has shown that the pupil diameter (PD) is inherently controlled by the Autonomic Nervous System (ANS) [GBSu][Hug95]. There is evidence that, in constant light conditions, the pupil diameter is increased when a subject is presented with stress stimuli. The reasons behind this phenomenon lies in a mechanism that modifies the balance between the Sympathetic and Parasympathetic divisions of the ANS [Hug95]. This effect needs to be taken for the development of the AffectiveMonitor (Figure. 4.1) for the evaluation of a computer users affective state based on the Circumplex Model of Affect [TOlR$^+$18].

It is known that pupil diameter changes are not only caused by affective reactions, but also by the amount of light that falls upon the retina, causing the Pupillary Light Reflex (PLR), which can be viewed as a process to regulate the amount of light reaching the retina [BW98]. This effect causes the contraction of the pupil and is superimposed to the changes in pupil diameter caused by affective responses, and, therefore, hinders our study. Thus, we seek to remove the PLR component from the pupil diameter signals we measure. Previous work from FIU DSP Lab research group [GBSu] has presented an approach that uses an Adaptive Interference Canceller (AIC) to remove the PLR component from the PD signal. That previous work utilized the AIC canceller to implement a stress detector tested on the reactions of the subject to Incongruent Stroop Segments. The study showed promising results in the PD-based systems performance as evaluated by the Receiver Operating Characteristic curve (ROC). The PD-based stress detector exhibited an area under the curve (AUROC) of 0.9331, indicating robust performance after the PLR was removed.

There are also other physiological signals that can act as indicators of arousal changes, such as the Galvanic Skin Response (GSR), the Blood Volume Pulse (BVP), the Heart Rate (HR). However, the pupil diameter is more suitable to estimate the arousal level and assess the affective state of a computer user because it can be observed non-intrusively, which is critical due to the nature of the study itself. In this kind of experiment, it is highly desirable that the subject remain at his/her normal state as much as possible, without any unnecessary distracting factors. This issue is also the reason why we chose to use the RGB camera from Kinect, which is already a part of the AffectiveMonitor system [TOlR$^+$18]], to measure the illumination around the subjects eye. Previously, a light meter was used to obtain illuminance signals to play the role of the required noise reference in the AIC algorithm. The light

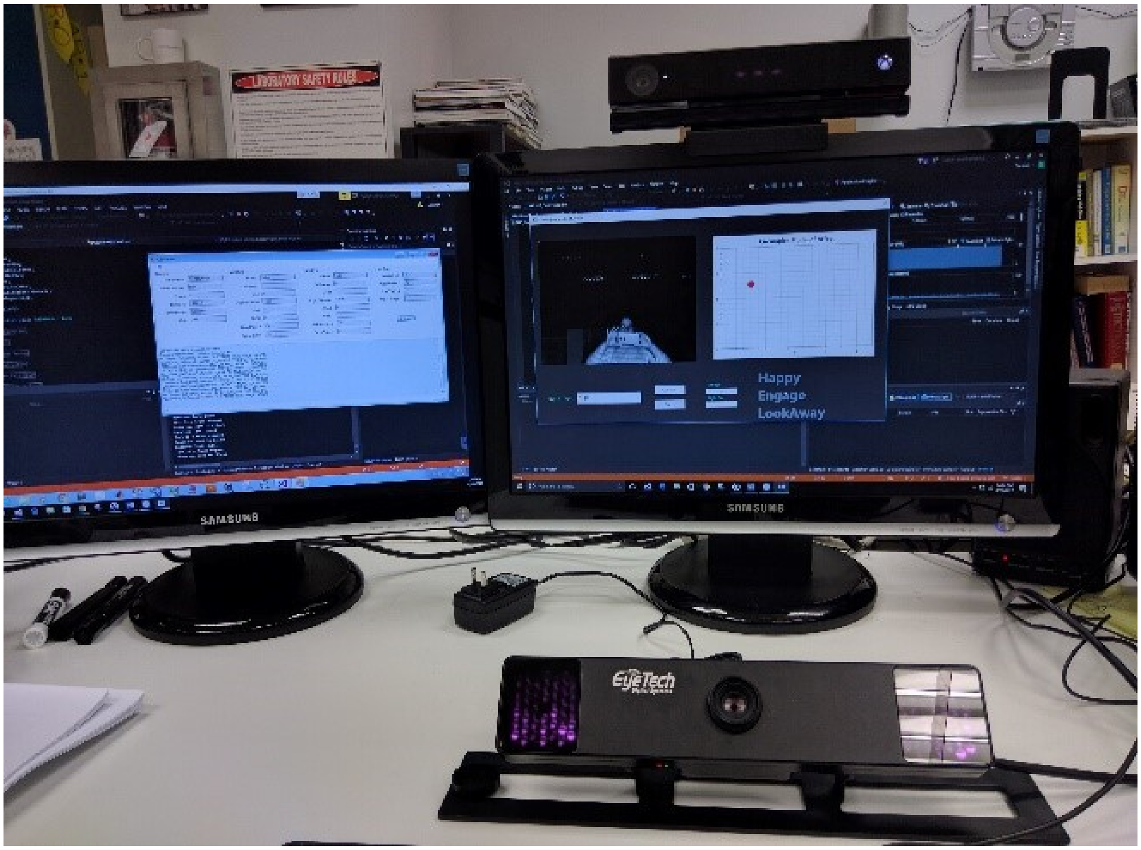


Figure 4.1: An entire system including Kinect V2 (on top of the screen) and TM3 (in front of the computer)

meter requires the placement of a sensor at the desired area where we would like to measure the illumination and it causes some distraction to the subject during the experiment.

In the following sections, we will discuss the AIC strategy in detail and describe how we obtain the illuminance signals around the eye area of the subjects face using images from the RGB camera as a mean to measure the illumination.

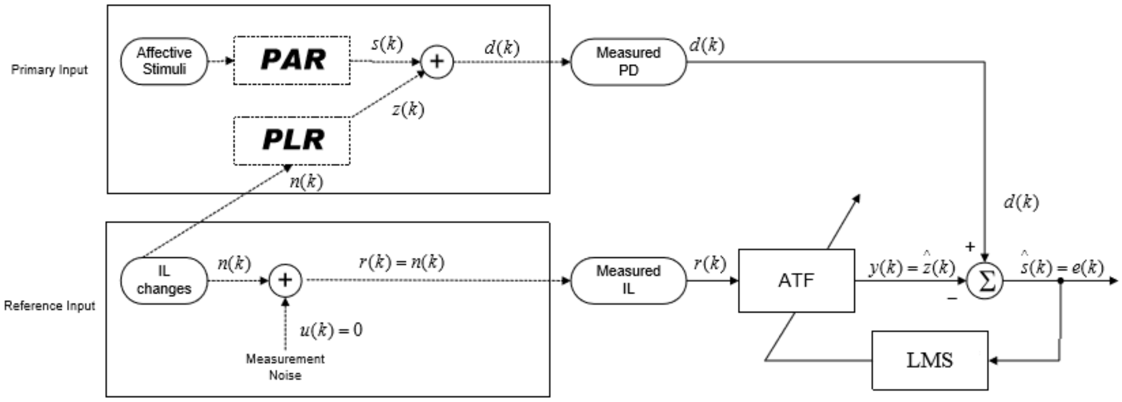


Figure 4.2: Diagram of Adaptive Interference Canceller (AIC) (from [GBSu])

4.2 Methodology

4.2.1 Adaptive Interference Canceller

The Adaptive Interference Canceller (AIC) is a system that is often used in Digital Signal Processing (DSP) to remove an unwanted interference component that pollutes a signal of interest [SK88]. The best way to explain how the system work is to walk through its diagram (Figure 4.2). The concept here is to measure the signal of interest $s(k)$ that is corrupted with an uncorrelated noise $z(k)$ as the primary input signal $d(k)$. The reference input signal $r(k)$ is a signal that is correlated with the corrupting noise $z(k)$ but uncorrelated with our target signal $s(k)$. The adaptive algorithm, the Least Mean Square (LMS), in this case, will adjust the parameters in an Adaptive Transversal Filter (ATF) to bring the reference input signal $r(k)$ to be as close as possible to the interference signal $z(k)$ in order to bring the error $e(k)$, down to a minimum value (in a mean squares sense). By doing so, we can obtain our signal of interest, i.e.,the filtered signal $\hat{s}(k)$, with the attenuated interference signal. In order to apply the theory to our application, we can think of the pupil diameter signal (PD) obtained from the TM3 Eye-Gaze Tracker as the primary input signal $d(k)$ while the measured illumination around the subject's eye area, from the RGB camera (Kinect) is used as the reference input signal $r(k)$. After the filtering process, we expect to obtain the output signal $e(k)$ that mainly contains the Pupillary Affective Response (PAR) component without the Pupillary Light reflex (PLR), which is removed by the adaptive filter.

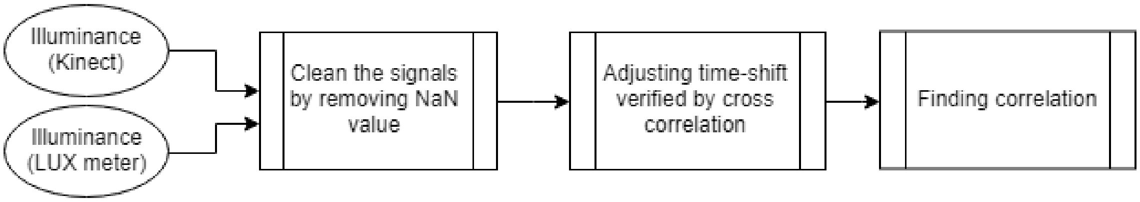

Illuminance (Kinect)
Illuminance (LUX meter)
Clean the signals by removing NaN value
Adjusting time-shift verified by cross correlation
Finding correlation

Figure 4.3: Diagram showing the process of finding correlation between Kinect and LUX meter signals

4.2.2 Kinect as LUX meter

As explained earlier, in the introduction section, the studies related to the evaluation of affective state require the subject to be in his/her normal condition as much as possible to minimize extraneous stimulation or distractions. We chose to utilize the RGB camera (Kinect) for an illumination measurement since Kinect is already a part in our system [TOlR$^+$18]. To explain the approach followed to achieve this, will first define a few terms used throughout the description of the process.

Luminance : Measured in candela per square meter is the parameter perceived by humans as the brightness of a light source.

RGB camera : Captures the incoming light rays and turns them into electrical signals enabling many pieces of electronic equipment to act as light detectors. The incoming color and brightness of an image are converted to numbers, preserving those characteristics of the image and breaking it up into millions of pixels, depending on the camera resolution.

For the purpose of eliminating the unwanted PLR factor from the pupil diameter signal using the RGB camera, we only compute the pixel values around the eye area, using a cropping rectangle image that always has its center between the left and the right eyes (Figure. 4.4).

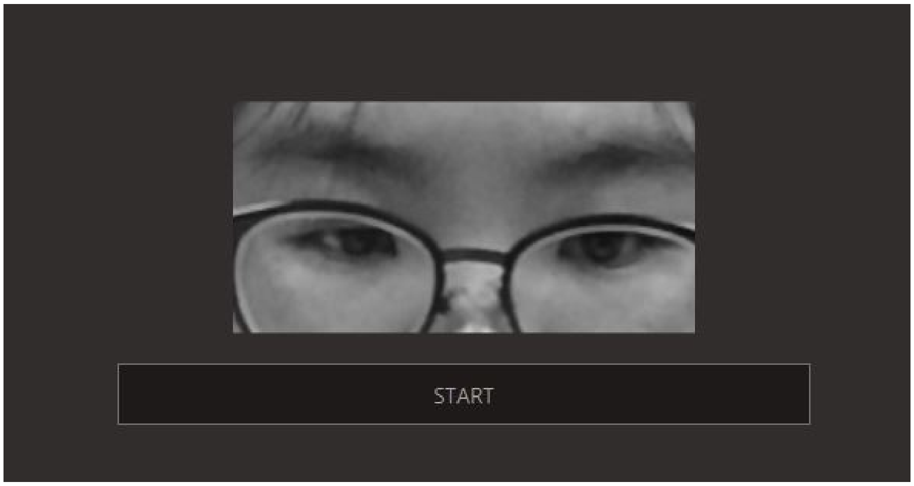


Figure 4.4: Cropped video used to compute luminance around the eye area

The pixel values in an image from the RGB camera are proportional to the luminance, because the light sensors convert the intensity of light falling upon them to electrical signals whose strength depends on the brightness of the received light. That is why an RGB camera can act as a luminance meter [HE11]. Equation 4.1 is used to calculate the luminance from RGB values in the image, according to a color model based on human physiological characteristics [AMu]. Note that, R is Red, B is Blue, and G is green.

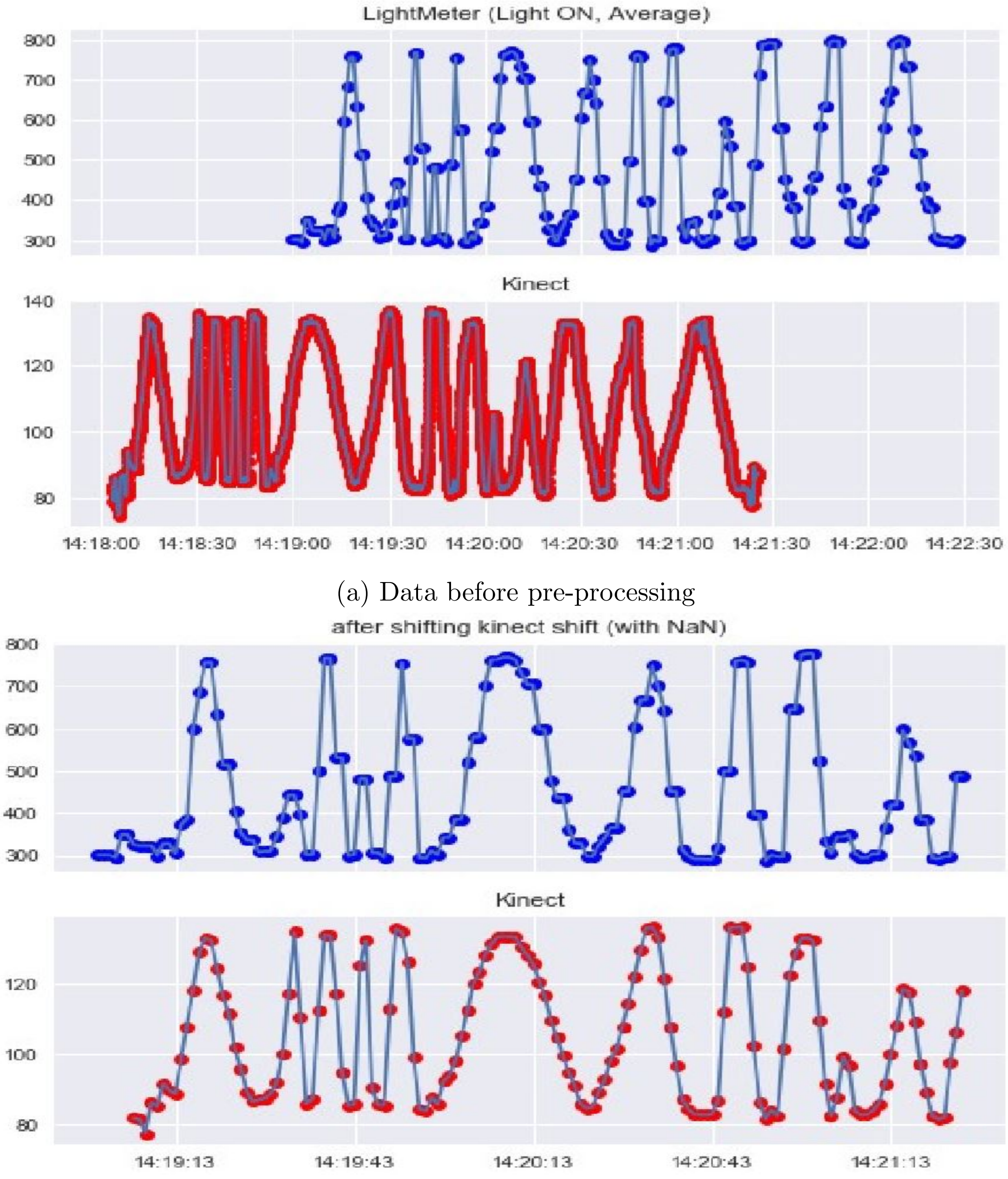
$$Y' = 0.299R' + 0.587G' + 0.114B' \tag{4.1}$$

However, the sensitivity of the sensors may be different for different RGB cameras. The relationship shown above may vary depending on the camera specifications. For this reason, we need to find out if the luminance values measured via our implementation followed the same trends as the luminance values measured using a luminance meter. To verify this hypothesis, we performed simultaneous light measurements in our experimental setup using the RGB camera from Kinect and a stand-alone LUX meter (Extech 401036 Datalogging Light Meter), while introducing strong illumination changes. Subsequently, after some processing, we computed the correlation between the two signals. If our hypothesis is correct, the luminance values obtained from Kinect should have a high correlation with the luminance value measured from the lux meter. A summary diagram of how we confirm our hypothesis is shown in Fig. 4.3

Table 4.1: Correlation Coefficient between Kinect and LUX meter

	Kinect	LUX meter
Kinect	1.00000	0.922234
LUX meter	0.922234	1.00000

We can notice that the plots in Fig. 4.5a are not synchronized because of the different delays in the measurement systems. To circumvent this problem, we performed a correlation analysis to determine the delay time and then re-align the two signals. After the aligning process, now we can calculate the correlation between the two signals. Figure 4.5b shows the plot of luminance after the preprocessing and shifting of one signal to align it with the other. Then we computed the correlation between these signals. The correlation coefficients of illuminance signals measured from Kinect and the LUX meter are shown in Table 4.1. The pairs of measurements are shown in a scatter plot in Fig. 4.6, which also includes a "best



(a) Data before pre-processing

(b) Data after pre-processing

Figure 4.5: Pre-processing of luminance signals obtained from LUX meter (blue) and Kinect (red)

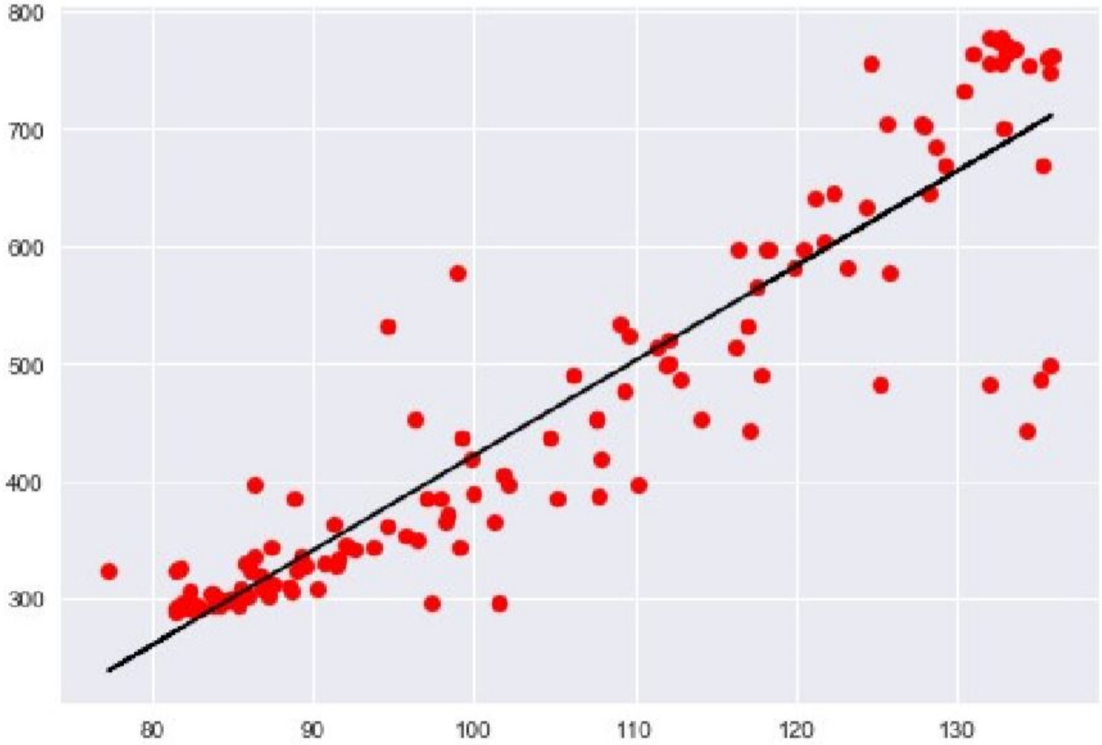


Figure 4.6: Scatter plot and correlation (m=8, b=-386)

fit” line. The result indicates strong correlation between the two signals, confirming that our hypothesis is correct and that we can use the luminance signal from our implementation as the reference input $r(k)$ (see Figure 4.2) to filter out the PLR from the pupil diameter measured signal.

4.2.3 Removing the Pupillary Light Reflex

As we have explained, we use an adaptive interference canceller (AIC) to filter out the Pupillary Light Response (PLR) and obtain a result, an output signal, containing only the Pupillary Affective Response (PAR). The first step is to pre-process the pupil diameter signal (PD); for instance, substituting missing samples due to eye blinks with an average value of the samples recorded in the neighborhood of the missing samples, and normalizing the pupil diameter as well as the illuminance signal before the filtering process. The signals we record are pupil diameter values obtained from the TM3 Eye-Gaze Tracking device from both left and right eyes containing about 7000 samples recorded at a sampling rate of 1 sample/sec. The illuminance signals are recorded using the RGB-camera (Kinect) at the same sampling rate. An example plot of pupil diameter signals after pre-processing along with the illuminance signal is shown in Figure 4.7.

The adaptive interference canceller (AIC) development follows the theory and practice from [TJ18] in order to implement an LMS adaptive filter. Both recorded pupil diameter signals that are impacted by the pupillary light response ($d(k)$) and the illuminance signal ($r(k)$) are normalized before they are processed by the LMS adaptive filter. There are two hyperparameters that affect the performance of the adaptive filter. They are the length of the delay line (L) and the learning rate (mu). The longer the delay line is, the slower and smoother the modified reference input signal ($y(k)$) becomes. In this case, we would like $y(k)$ to imitate the PLR component in the primary input ($d(k)$) as much as possible so the output signal ($e(k)$) is only left with the PAR after $y(k)$ is subtracted from $d(k)$. The learning rate (mu) determines how fast the filter can adapt to its target. Setting the right learning rate (mu) is critical here since if it is set too low, the filter could have a degraded performance; while the system might be unstable if it is set too high. In our study, we set the delay line length (L) at 10 and the learning rate at 50. We will discuss our results in the next section.

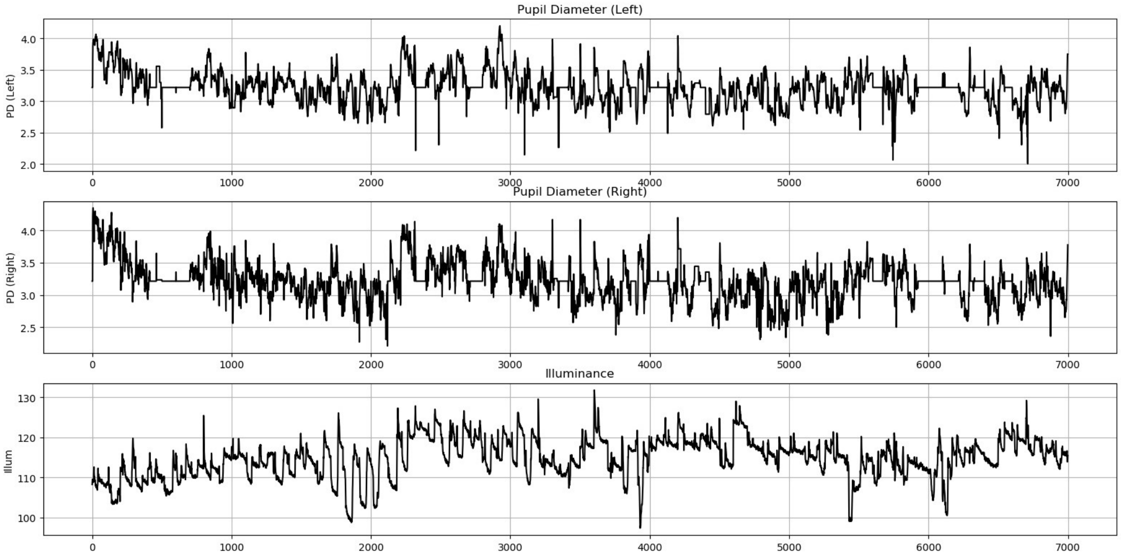


Figure 4.7: Plots from top to bottom: Pupil Diameter (Left), Pupil Diameter(Right), Illuminance

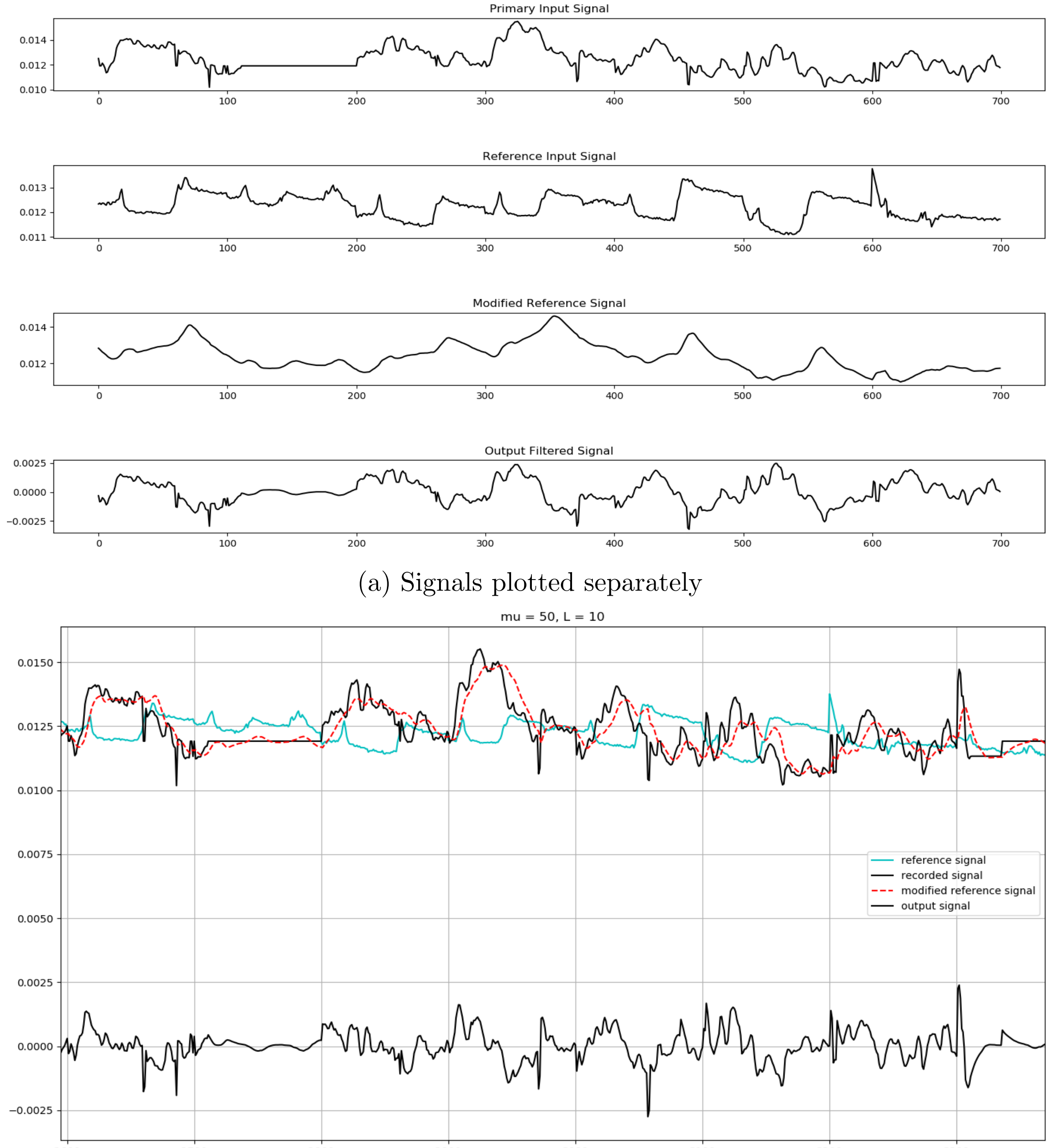
4.3 Result

An example of the results obtained with the AIC is shown Figure 4.8, which consists of two plots. The first one (Figure 4.8a) shows signals $d(k), r(k), y(k)$, and $e(k)$, respectively, from top panel to bottom panel. Figure 4.8b shows some of these same signals superimposed for an easier visualization. Here each signal is shown in a different line style. The primary input signal ($d(k)$) is represented in solid black line at the top part of the graph; this signal is the left pupil diameter signal. Our output signal ($e(k)$) is also in solid black line but located at the bottom of the graph. The reference signal ($r(k)$), illuminance, is shown here in light color and, lastly, the modified reference signal ($y(k)$) is plotted in dotted line. It is possible to observe, in this figure, how each signal behaves based on the nature of the pupillary response. The basic idea is that when the illuminance is low, the pupil diameter is increased in order to adjust the amount of light that reaches the retina. That is why when the reference signal is lower, the pupil diameter signal is shifted upward.

The purpose of the implementation of the adaptive filter is to eliminate this effect and shift the pupil diameter down to the baseline when the interference from pupillary light response produced by illumination changes occur. In Figure 4.8b, we observe that the output signal behaves as we expected. In these instances, the output signal in the plot did shift down to the base line while still preserving the PAR information. Hence, the LMS filter seems to be removing the influence of the pupillary light response from the pupil diameter signal.

In this chapter, we described the processing of images from an RGB-camera for substitution of a LUX meter to measure the illuminance around the eyes of the subject. Our testing showed a correlation of 0.922 between the illuminance signals obtained by the LUX meter and the Kinect camera.

We then used the illuminance signal obtained through the RGB camera in Kinect as the noise reference signal in an adaptive interference canceller. In the canceller architecture the measured pupil diameter signal is the primary input and comprises both, the pupillary light response (interference) and the pupillary affective response (signal of interest). The behavior of the results obtained from the adaptive interference canceller seem to indicate that canceller is, in effect, compensating for pupil diameter signal shifts that are clearly occurring in response to illumination changes.



(a) Signals plotted separately

(b) Signals plotted in one graph

Figure 4.8: Plot of signals processed in removing PLR using AIC

CHAPTER 5

THE AFFECTIVE ASSESSMENT SYSTEM

This chapter describes the implementation of the affective assessment system developed through this research, which has been named AffectiveMonitor. This is the system used to visualize, collect, and analyze the data used in this study. Chapter 2 explained that we target the estimation of two parameters (arousal and valence) to assess the affective state of the user based on the Circumplex Model of Affect introduced by Russell. For this purpose, the system collects two types of data having a direct relationship to arousal and valence, which are the pupillary response and the facial expression, respectively. These data are collected in order to train a predictive model that will identify the users affective state while he/she is interacting with a computer. For those purposes, the system developed seeks to acquire the facial expression and the pupil diameter data in a natural way that will not interfere or distract the user during his /her ordinary interaction with a computer. The AffectiveMonitor system, described in the following sections, was developed with those considerations in mind. The system consists of three elements which are a Kinect module, a TM3 Eye-gaze Tracking Device, and the AffectiveMonitor Software Application.

5.1 Kinect

Kinect is a product developed by Microsoft as a result of their interest in creating an alternative type of game controller for their Xbox console game station via a natural user interface. We can consider Kinect, basically, as a motion sensing input device. Beyond its intended gaming application, Kinect has also attracted significant attention from the research and development community. This low=price, powerful sensor device has opened new opportunities for human-computer interaction research. Microsoft has produced a couple of generations of Kinect including the first generation Kinect 360, Kinect for Window, Kinect for Xbox One, and Azure Kinect. In our study, we chose to work with Kinect for Xbox One which is the third generation of hardware and software for the gaming console. It includes a color video camera (RGB), a time-of-flight (TOF) depth sensor, an Infrared Camera, a Microphone array, and the corresponding Software Development Kit (SDK). The combination of an RGB camera and a depth sensor, which is referred to as an RGB-D camera, yield both a color image and a depth map, which are used to generate a 3-dimensional map of Kinects vicinity. In particular, the HD face framework, available from the SDK, can generate a 3-dimensional representation of objects of interest, e.g., is the face of a computer user, in a meshes object. Furthermore, Kinect also supports the implementation of a facial recognition task, which means that after the users face has been scanned properly, Kinect will generate a meshes object that has been tailored to more specifically represent the individual users face shape. The meshes object is the collection of many triangles that are structured and aligned together to correspond with the shape of the 3-dimensional object. Each triangle has a position based on the locations of its vertices, which are part of the 3-dimensional points cloud collected by the RGB-D camera. We can yield the 3-dimensional points cloud of the meshes object of the users face from HD face framework as well. There are 1374 3-dimensional points that represent the facial meshes object as shown in Figure 5.1

According to the limited documentation provided by Microsoft, these 3-dimensional points are stored and indexed in a specific order, and only a selected list of indices that corresponds to the critical anatomical landmarks is published. (The list of indices and their description are shown in Table 5.1.)

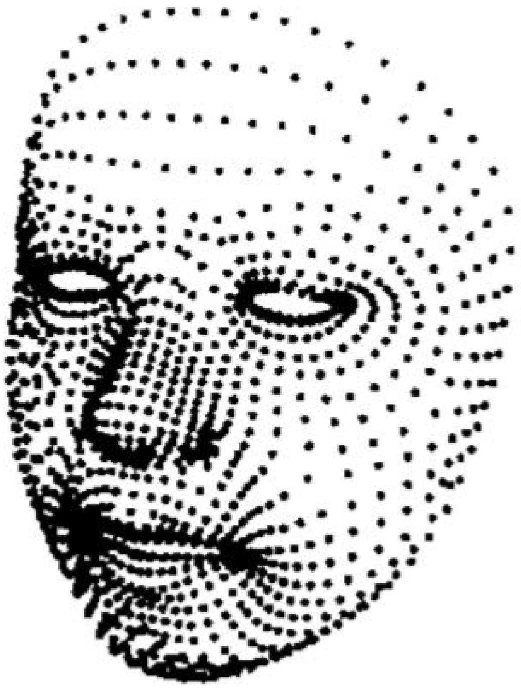


Figure 5.1: HD Face vertices from a detected face [Rah17]

Although the HD face framework provides the developer with vital building blocks, the framework only provides the necessary features just for the developer to get started in developing an actual project. We have integrated the HD face framework to AffectiveMonitor software application to benefit from all that the Kinects framework can offer.

5.2 TM3 Eye-gaze Tracking Device

The EyeGaze Tracker used in this study is the TM3 EyeGaze Tracker (TM3 EGT) device from EyeTech Digital Systems Inc. This is a compact desktop EGT system, suitable for the monitoring of a computer user seated at a desk. The system consists of a high definition camera, infrared sources, and its software environment. It is capable of tracking both eyes in real time and it can be used with any windows-

Table 5.1: List of indexes of 3-dimensional points and their description provided by HD face framework

Key	**Index**
HighDetailFacePoints_LefteyeInnercorner	210
HighDetailFacePoints_LefteyeOutercorner	469
HighDetailFacePoints_LefteyeMidtop	241
HighDetailFacePoints_LefteyeMidbottom	1104
HighDetailFacePoints_RighteyeInnercorner	843
HighDetailFacePoints_RighteyeOutercorner	1117
HighDetailFacePoints_RighteyeMidtop	731
HighDetailFacePoints_RighteyeMidbottom	1090
HighDetailFacePoints_LefteyebrowInner	346
HighDetailFacePoints_LefteyebrowOuter	140
HighDetailFacePoints_LefteyebrowCenter	222
HighDetailFacePoints_RighteyebrowInner	803
HighDetailFacePoints_RighteyebrowOuter	758
HighDetailFacePoints_RighteyebrowCenter	849
HighDetailFacePoints_MouthLeftcorner	91
HighDetailFacePoints_MouthRightcorner	687
HighDetailFacePoints_MouthUpperlipMidtop	19
HighDetailFacePoints_MouthUpperlipMidbottom	1072
HighDetailFacePoints_MouthLowerlipMidtop	10
HighDetailFacePoints_MouthLowerlipMidbottom	8
HighDetailFacePoints_NoseTip	18
HighDetailFacePoints_NoseBottom	14
HighDetailFacePoints_NoseBottomleft	156
HighDetailFacePoints_NoseBottomright	783
HighDetailFacePoints_NoseTop	24
HighDetailFacePoints_NoseTopleft	151
HighDetailFacePoints_NoseTopright	772
HighDetailFacePoints_ForeheadCenter	28
HighDetailFacePoints_LeftcheekCenter	412
HighDetailFacePoints_RightcheekCenter	933
HighDetailFacePoints_Leftcheekbone	458
HighDetailFacePoints_Rightcheekbone	674
HighDetailFacePoints_ChinCenter	4
HighDetailFacePoints_LowerjawLeftend	1307
HighDetailFacePoints_LowerjawRightend	1327

based communication software. More details regarding the technical specification of the device can be found in [COG11].

In this study, we used the Software Development Kit provided by EyeTech Digital, called Quicklink2 as a base to develop the secondary side of AffectiveMonitor. Quicklink2 is written in C++ and builds upon the OpenCV library, which is a popular C++ library for computer vision and image processing applications. But since our AffectiveMonitor system is written in the C# language, we preferred to develop the secondary side software of AffectiveMonitor to be written in C# as well. Fortunately, there is a Quicklink2 Microsoft .NET wrapper written in C# available to use as an open-source for the developer community. This wrapper was written by Justin Weaver. The author has provided all the necessary information for running the application in his website [Jus09]. We have used his wrapper in developing the secondary side of AffectiveMonitor, which includes a number of custom features, such as sending pupil diameter samples to the primary side.

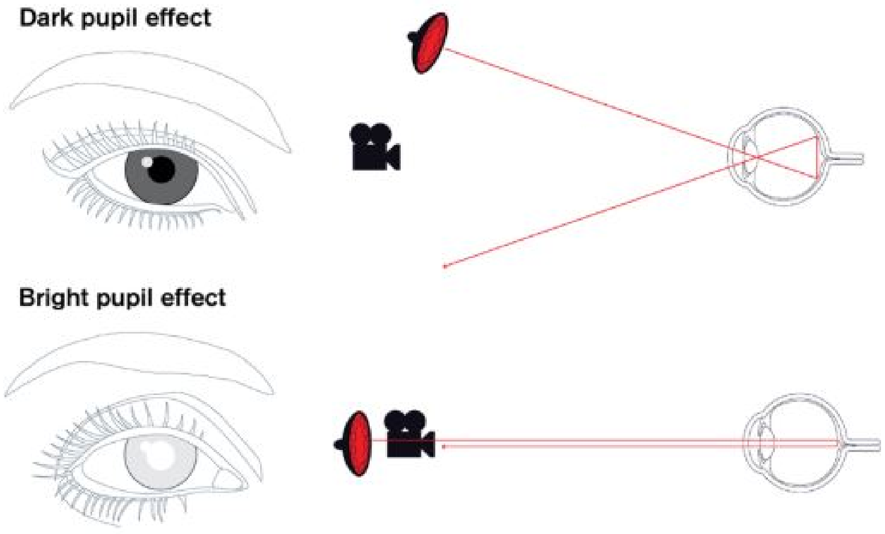


Figure 5.2: Difference of the illuminator's placement between bright pupil effect and dark pupil effect (Picture from [Tob15])

TTM3 EGT uses the Dark pupil effect to track the pupil diameter center and the corneal reflection. There are two main approaches for identifying and tracking the center of the pupil in infrared video-based EGT systems. In the bright pupil effect method, an illuminator is located near the optical axis of the camera, which causes the pupil to appear brighter than the rest of the video frame. In contrast, the dark pupil effect uses an illuminator placed away from the optical axis of the camera, which results in the pupil having a darker shade than the rest of the frame (See Figure 5.2). Both methods have been used successfully by the different manufacturers of eye gaze tracking equipment.

5.3 AffectiveMonitor Software Application

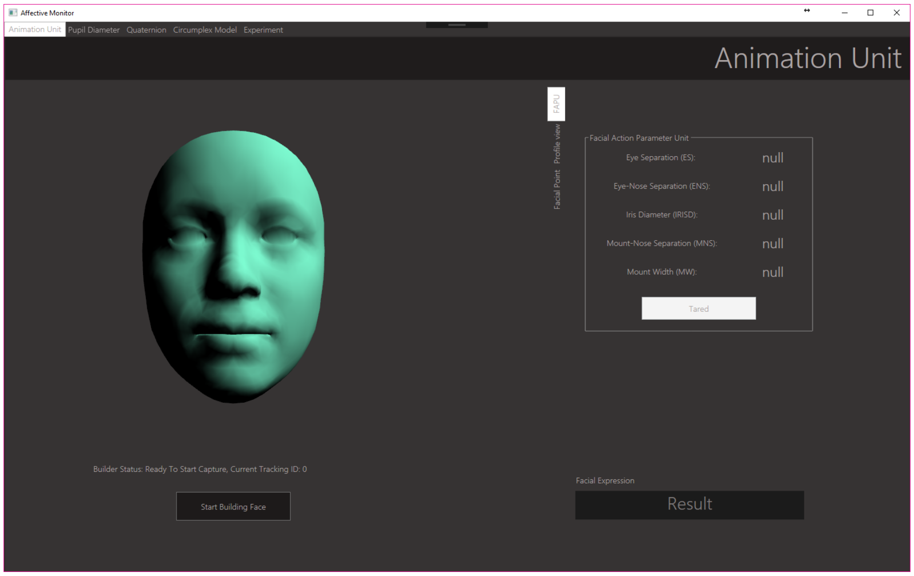
The idea pursued in this project, namely to combine two types of observed data, the pupil diameter (arousal) and embedded information in the 3-D facial expression (valence), to assess the affective state of a computers user is a novel approach to affective assessment. Accordingly, there are not many available off-the-shelf digital tools for researchers in this field. Some tools are available separately and offer challenges in their integration. In order to circumvent these problems, an integrated custom software application was developed for this research, that connects with both Kinect and TM3 EGT device and is useful to collect and to visualize the obtained data. We called this software AffectiveMonitor. It consists of two parts, a part that connects with Kinect (which we call the Primary Side), and another that connects with TM3 EGT device (Secondary Side).

5.3.1 Primary Side

The primary side of the AffectiveMonitor interfaces with the Kinect module. This part is responsible for translating the data from the hardware device (Kinect in this case) into files that can be used for the development of the predictive model and its application. It is also responsible to provide the investigators with the ability to interact with the data captured by Kinect. It makes use of the software framework (libraries) provided by the Kinects developer team. These have been integrated into the AffectiveMonitor system to do designated tasks. Other responsibilities of the primary side are the embedding of the the facial expression, the reception of the pupil diameter data from the secondary side of AffectiveMonitor, storage of the data in appropriate formats, measurement of the illumination around the eyes of a computers user, data visualization for verification purposes, and the step-by-step implementation of the protocol designed for the experimenting sessions. Figure 5.3 shows the first appearance of the AffectiveMonitor (Primary side). Please notice that there are tabs located at the top of the window. They allow the use of each one of the modes of the AffectiveMonitor. The following subsections will explain in detail, the uses and the functionalities of each mode in the AffectiveMonitor application.

Home

The best way to describe how AffectiveMonitor works might be to describe the functionality of the components presenting in each mode along with their layout pictures. Figure 5.3 shows the default mode of AffectiveMonitor. The purpose of this page is to show the interaction with Kinect. This page includes visualizing the detected face and its meshes object, collecting Facial Animation Parameters (FAPs), observing animation units provided by the Kinect library framework, and displaying the facial cloud of points of the face object. On the left side, it shows a 3-dimensional facial mesh, which we will refer to as the face model in the rest of this dissertation. Initially, space is empty, before a face has been detected.



Facial Action Parameter Unit

Eye Separation (ES): null

Eye-Nose Separation (ENS): null

Iris Diameter (IRISD): null

Mount-Nose Separation (MNS): null

Mount Width (MW): null

Builder Status: Ready To Start Capture, Current Tracking ID: 0

Figure 5.3: AffectiveMonitor in its default tab which is the first page to show when AffectiveMonitor is first opened. On the left side shows the 3-dimensional face model in its initial state. On the right side is the FAP info panel that will display the FAP unit values collected after the FAP unit are measured.

Everything starts when Kinect finds the presence of the users face in front of its sensor. From there, it will start representing the movement of the facial muscles of the user, as detected, to the face model. This results in a mirroring behavior between the users face and the face model displayed in the application. Kinect also provides the feature where we can adjust the face model to morph more specifically to the face shape of the user it is detecting. This process requires the scan of the users face in order to achieve the enhanced similarity in the face model. We have integrated this feature to increase the accuracy of the data that we to collect. The scanning process can be activated after the Start Building Face button is clicked. First, the user is asked to look straight forward, then he/she is asked to tilt his/her face to the right, tilt his/her face to the left, and finally tilt his/her face up to complete the scanning process. The text area below the face model is used to indicate the status of the face model process. It will display one of the following messages:

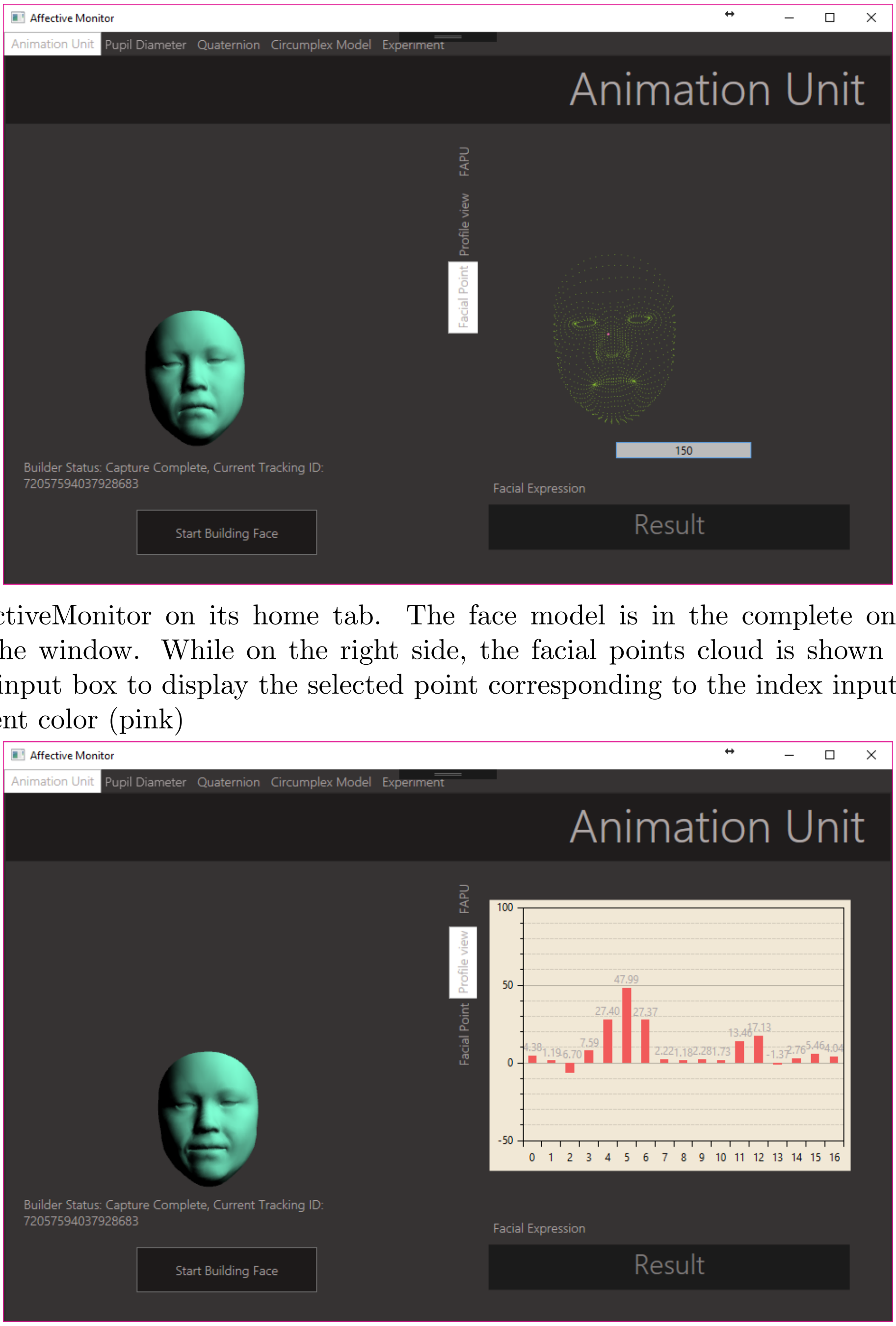
(i) Initial state (No face detected)

(ii) Default state (Face detected, Before calibration)

(iii) Building state (Face detected, Under calibration)

(iv) Complete state (Face detected, After calibration)

On the right side of the window, one of three display modes can be. The first one is the Facial Animation Parameter Unit (FAPU) which we have explained in Chapter 3. We can obtain FAPUs of individual users by clicking the Tared button below the FAPU panel on the right side of the window. On the moment the button is clicked, the FAPU parameters will be captured so before we collect FAPUs, we have to ask the user to make a neutral face. The second one is the Animation Units (AUs). AUs are calculated based on the changes in the shape of the face model [Rah17]. AUs are provided by the Kinect framework and they are used in animation production as a convenient way of animators to map the captured facial expressions of a person to the facial expression of their 3-dimensional animated characters. We display AUs in dynamic bar charts as shown in Figure 5.4b on the right side of the window. This way of visualizing the data enables us to gain some insight into observing the typical bar graph profiles that accompany specific facial expressions displayed by the user. AUs are determined based on the principles of the Facial Action Coding System (FACS) [EFA80] and they have 16 units in total. Lastly, the third form of information we yield from Kinect is the actual facial point cloud, which is just an alternative way to display the face object. This can be seen in Figure 5.4a, where the face model is displayed as points, instead of a mesh, on the right side of the window. For this research, we also implemented a custom way to access the index of Kinect facial points. The default color of the points is lime green. However, in this custom display, a specific index may be typed in the input textbox (below the face



Builder Status: Capture Complete, Current Tracking ID: 72057594037928683

(a) AffectiveMonitor on its home tab. The face model is in the complete on the left side of the window. While on the right side, the facial points cloud is shown with the number input box to display the selected point corresponding to the index input number in different color (pink)

Builder Status: Capture Complete, Current Tracking ID: 72057594037928683

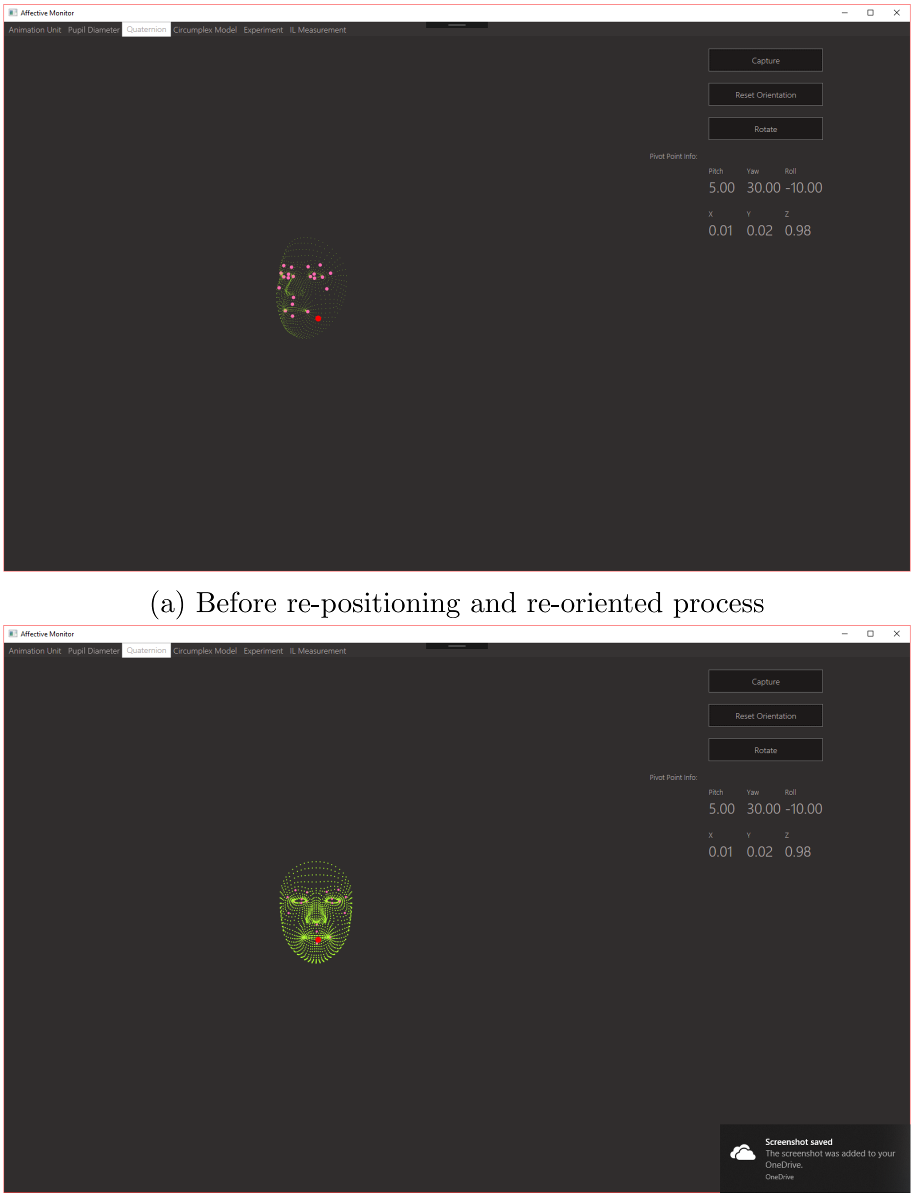
(b) AffectiveMonitor on its home tab. The face model is in the complete state after the calibration process. On the right side of the window, shows the plot of animation units (AU) provided by Kinect library framework

Figure 5.4: AffectiveMonitor: Kinect Interactive Mode

points model), and the corresponding specific landmark point will appear in pink. We created this because we needed a way to verify that the indices that we select to observe are at the right location on the face model (since the Kinect developers provide a very limited amount of documentation). Finally, on the bottom right of the window is where the result of our predictive model will be displayed.

Position and Orientation

An obstacle that emerged in trying to embed the facial point cloud to FAP vectors, was that the position and the orientation of the face model are not constant, because the subject may move his/her face during recording. Fortunately, Kinect doesnt track only the changes in facial muscles, but also the position and the orientation of the users face. The FAP vectors are measured in a fixed direction between two selected points either in the x-axis, y-axis, or z-axis. Therefore, it was necessary to develop a solution to account for the fact that in our position and the orientation of the face model could be different in every frame. The approach we used to solve this problem is to always re-orient the face model to be in a neutral position (orientation) and translate the face model to the origin before we calculate the FAP vectors. We describe how we apply the Quaternion Rotation Operation to rotate the face model and translate the face model back to the desired position in Chapter 3. To verify if our implementation is correctly rotating and translating the face model, we have included the Position and Orientation mode in AffectiveMonitor. The purpose of this tab is to enable us to observe the face point model that we captured before and after the re-orienting and re-positioning process. Figure 5.5 5 demonstrates the Quaternion tab which includes the window to display the face point model on the left side. The right side is where all the buttons to apply the process are located. The first button on top is labeled Capture whose name implies its functionality.



Pivot Point Info:

Pitch	Yaw	Roll
5.00	30.00	-10.00

X	Y	Z
0.01	0.02	0.98

(a) Before re-positioning and re-oriented process

Screenshot saved
The screenshot was added to your OneDrive.
OneDrive

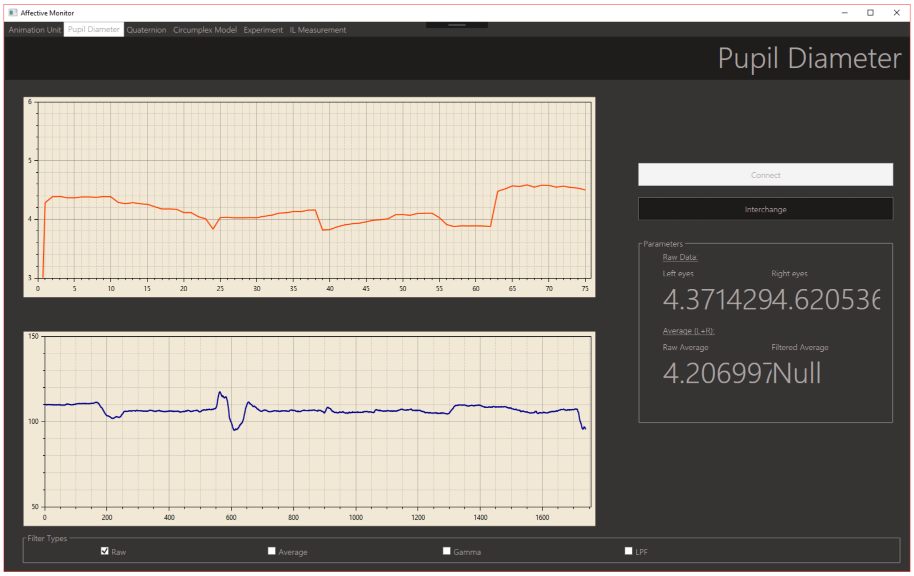
(b) After re-positioning and re-oriented process

Figure 5.5: AffectiveMonitor on its quaternion tab. The face points model of one frame is captured after the ”Capture” button is clicked. Also the angle (in euler) and the position (cartesian domain) of the face points model are displayed on the right side of the window. The point in red represents the pivot point of the model. Those in pink color are the selected points for obtaining FAP vector

When the Capture button is clicked, the face point model of one frame is captured and displayed on the canvas on the left side of the window. The second button Reset Orientation and Rotate are for applying the re-orienting and re-positioning processes to the face points model. After clicking the Rotate button, the position and the orientation of the face points model are reset to the neutral position. The default color of the face points model is in lime-green color. Those that are in pink color are the points that will be used to calculate the FAP vector. There is also one red point that is displayed in the canvas. We call this point in red a pivot point. This point is an important point to apply the re-positioning and re-orienting processes because it contains information about the position and the orientation of the face point model. The pivot point is determined by the centroid of the face model. The other points in the model are referenced to this point. This means that, if the pivot point is moved to another position, all the points in the face model can be reconstructed by referring the pivot point as the origin. We also display the position (cartesian domain) and the orientation (in Euler unit) below the set of buttons for debugging purposes.

Pupil Diameter

Another parameter that we monitor in order to estimate the arousal level of the user is the pupil diameter signal. In our system, pupil diameter signals are received from the secondary side of AffectiveMonitor application (which will be discussed in the latter part of this chapter) and we would like to observe the signal in real-time. That is why this pupil diameter mode was created. Here, we implemented a dynamic plot of the pupil diameter signal to monitor its behavior as it correlates to the changes in arousal of the user. On the right side of the window placed the buttons used to connect the primary side and the secondary side of AffectiveMonitor together via



Parameters

Raw Data:

Left eyes Right eyes

4.3714294.62053€

Average (L+R):

Raw Average Filtered Average

4.206997Null

Filter Types

☑ Raw □ Average □ Gamma □ LPF

Figure 5.6: AffectiveMonitor in its pupil diameter tab. The first row of plots shows the pupil diameter signals received from AffectiveMonitor (Secondary side). The bottom row plot shows the illumination signal obtained from RGB-camera. On the bottom right of the window shows the panel displaying the pupil diameters of left and right eyes as well as the average between two eyes.

the TCP/IP protocol. To observe the pupil diameter signals both sides of Affcetive Monitor need to communicate. The primary side can send the connection request clicking the button Connect on the top right of the window. If the connection is established successfully, the pupil diameter signals that are being received every 33 milliseconds will be plotted in the top panel of the graph. The Interchange button is used to send a single request asking the secondary side to send one sample of pupil diameter back to the primary side. This feature is very useful for checking the establishment of the connection and for debugging. The bottom panel is used to plots the illumination signal (This will be discussed later, in the explanation of the in Illumination tab). The plot of the illumination signal was placed here because we want to observe how the pupil diameter of the user responds to the changes in illumination signal as we already know that the pupillary light reflex (PLR) is one of the factors that affect the pupillary response. These plots are real-time dynamic plots that are very useful to observe the behavior of the pupil diameter signal.

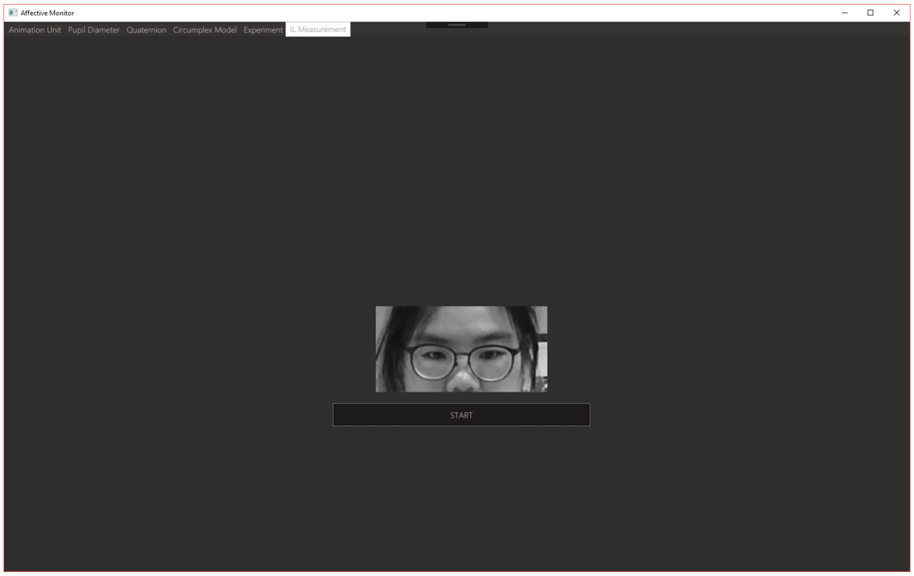


Figure 5.7: AffectiveMonitor in its illumination tab. At the center of the window shows the video of the are of interest (around a user's eye) where the illuminance value will be calculated based on this video obtained by Kinect's RGB-camera.

Illumination

As we mentioned earlier that Pupillary Light Reflex (PLR) is one of the factors that influence the changes in the pupil diameter signals. Changes in the amount of light reaching the subjects retinas will tend to introduce pupil diameter changes that are separate from those induced by affective responses. Even if the ambient illumination is kept constant, the brightness of different images displayed in the computer system can introduce those PLR responses. Previous research has used an actual light sensor, attached to the subject near his /her eyes, to measure the level of illumination present, in order to attempt to compensate for its effects. However, this unnatural fact will directly affect the users mood in one way or another. To avoid this problem, we chose to utilize an RGB-camera (as the one included in Kinect) to obtain an indirect measurement of the illumination around the subjects eyes. The details of the approach followed to accomplish that goal are explained in chapter 4. In this tab, AffectiveMonitor displays the video of the area of interest (around the users eyes) in grayscale (black and white). The button START below the video is clicked to start the illumination signal measuring process. The plot of the signals is shown in the pupil diameter tab for the monitoring purposes.

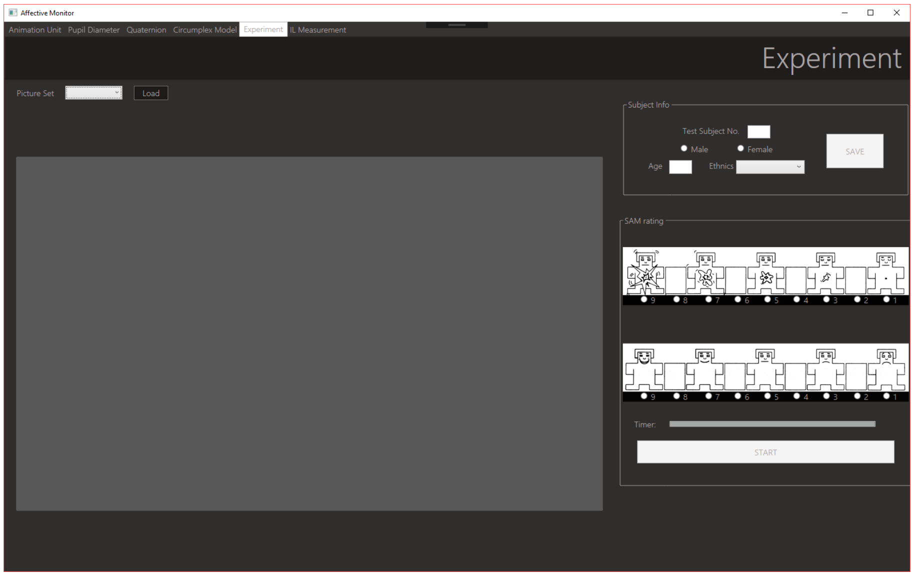


Figure 5.8: AffectiveMonitor in its experiment tab.

Experiment

In experiment tab (Figure 5.8), is used to initiate the execution that drives the experimental protocol designed for this research, as described in Chapter 3. On the left side of the window, the IAPS pictures are shown. We can choose the set of IAPS pictures that will be displayed sequentially by clicking on the dropdown menu and selecting the desired set of pictures, followed by a click of the Load button to load all the pictures. If none of the picture sets is selected, the default set will be loaded. The top right of the window is where the information from the test subject is collected. The Subject Info panel includes the test subjects number, gender, age, and ethnicity (Notice that the subjects name is not recorded). After the SAVE button is clicked, the program will automatically generate the folder whose name corresponds with the test subjects number and all the files that will be recorded. Below the Subject Info panel is where the SAM ratings are entered by the subject. Here, the user is asked to rate their arousal and valence in response to each IAPS picture displayed, using the SAM tool, immediately after each picture is displayed. The START button is clicked to start the experiment. When the experiment starts, 70 pictures are shown one after one with a gray screen in between to provide a resting period.

5.3.2 Secondary Side

The secondary side of the software application of AffectiveMonitor system has similar purposes as the primary side, but it is focused on the acquisition and transmission of the data from the TM3 Eye Gaze Tracker. The secondary side has three main applications that we use in our study. First, it can display the video captured from the TM3 EGT devices high-definition infrared camera so we can adjust the position and the angle of the camera to aim the field of view of the camera at the area around the eyes of the user. Second, it translates the pupil diameter signals from the TM3 EGT device into the format that is appropriate for transmission to the primary side and for storage. Lastly, it sends the pupil diameter signal as per request from the AffectiveMonitor primary side. The following subsections describe the software implementation of the AffectiveMonitor secondary side, which has two modes of operation: Video capture and EyeGazeInfo capture.

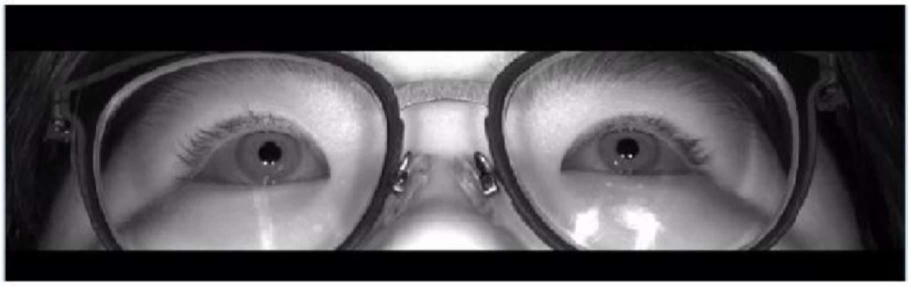


Figure 5.9: AffectiveMonitor secondary side on Video capture application.

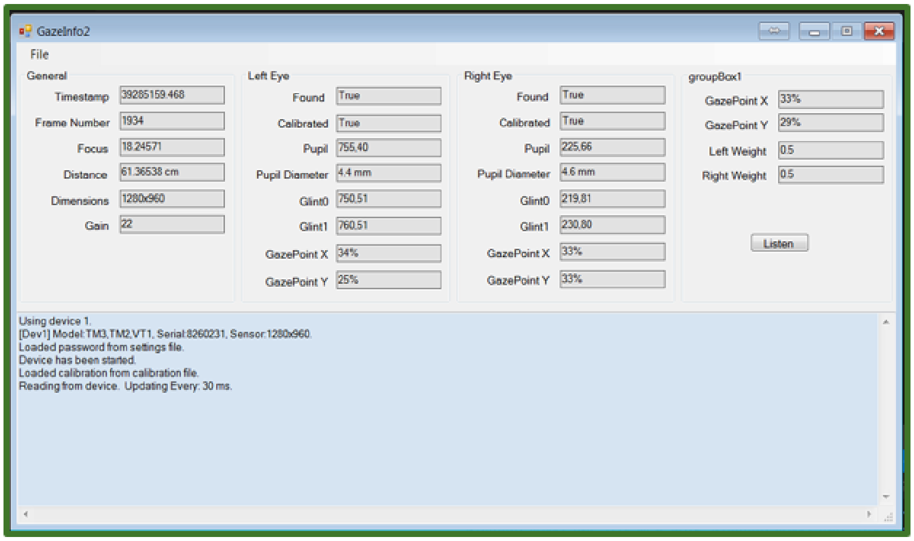
Video Capture

In this application, the window is just a simple video player where the view of the camera from the TM3 EGT is shown. The position and the angle of the camera have to be set properly in order for the device to detect the pupil diameter. Another factor that affects the pupil diameter detection application is the focus of the camera which we can adjust on the camera itself. First, the video will display the whole face of the user. Then, after the position, angle, and focus are adjusted properly, the device will display only the cropped video of the area around the eyes of the user, indicating that the pupil diameter is detected. Figure 5.9 demonstrates the video after the proper adjustment is done successfully.

EyeGazeInfo Capture

After we are certain that the camera is adjusted properly, the EyeGazeInfo capture application can be started. Figure 5.10 portrays the window panel of the application. The panel at the top part of the window exhibits all the information that TM3 EGT captures from the users eye; nonetheless, the only parameters that will be transmitted to the primary side are the pupil diameters from both left and right eyes. The bottom panel is where the log of the status of the application is shown.

To start receiving the request from the primary side of the AffectiveMonitor, we can click the Listen button at bottom-right in the top panel of the window.



General

Timestamp: 39285159.468
Frame Number: 1934
Focus: 18.24571
Distance: 61.36538 cm
Dimensions: 1280x960
Gain: 22

Left Eye

Found: True
Calibrated: True
Pupil: 755,40
Pupil Diameter: 4.4 mm
Glint0: 750,51
Glint1: 760,51
GazePoint X: 34%
GazePoint Y: 25%

Right Eye

Found: True
Calibrated: True
Pupil: 225,66
Pupil Diameter: 4.6 mm
Glint0: 219,81
Glint1: 230,80
GazePoint X: 33%
GazePoint Y: 33%

groupBox1

GazePoint X: 33%
GazePoint Y: 29%
Left Weight: 0.5
Right Weight: 0.5

```
Using device 1.
[Dev1] Model:TM3,TM2,VT1, Serial:8260231, Sensor:1280x960.
Loaded password from settings file.
Device has been started.
Loaded calibration from calibration file.
Reading from device.  Updating Every: 30 ms.
```

Figure 5.10: AffectiveMonitor secondary side on EyeGazeInfo capture application.

5.4 Summary

AffectiveMonitor is the custom data acquisition and processing system developed for this research. The system is suited to interact with the data capture from a Kinect module and the TM3 EGT. There are three elements in the system: Facial expressions capture, pupil diameter signal capture, and data interactive application. The reason behind our decision to separately develop the primary side and the secondary side instead of one complete application running in a single computer is for keeping the independence between these two devices. Development of the software that controls each device separately can help in dealing with conflicts in operating systems versions and port types required by the individual devices. For example, in our case, Kinect requires a USB3 port and Window 10 in order to run efficiently; while the TM3 EGT device runs in Windows 7 and needs a firewire port to connect to a computer.

CHAPTER 6

FEATURE EXTRACTION

It is known that the success of a predictive model depends to a large extent on the use of distinctive and meaningful features as its inputs. Therefore, the process of extracting those meaningful features from the signals collected for this research is a critical stage of its development. In order to ease the training of a predictive model and achieve a good result in accuracy, the feature extraction process is really essential. It plays an important role to present the classifier model with useful features. This helps accelerate the training process and eases the problem of having a limited amount of data.

In this research, the two types of data involved are pupil diameter signal and facial action parameter vectors as described in Chapter 3. This chapter discusses in detail the pre-processing and feature extraction performed on each one of these types of data.

6.1 Pupil Diameter signal

At first glance, the pupil diameter signal records that we obtained all have similar trends that can be observed by looking at the overlapping plot of all the 70 PD signals obtained from one test subject (for example, in Figure 6.1). A first peak is apparent during the first 2 seconds of a typical record. Then the signal descends rapidly, until it reaches a minimum point, and then bounces back, approaching an approximately steady level.

The behavior of the signal can be explained better by taking several factors into consideration. Refer to the diagram in Figure 6.4, which indicates that a resting screen in gray is shown before the presentation before each IAPS picture is shown.

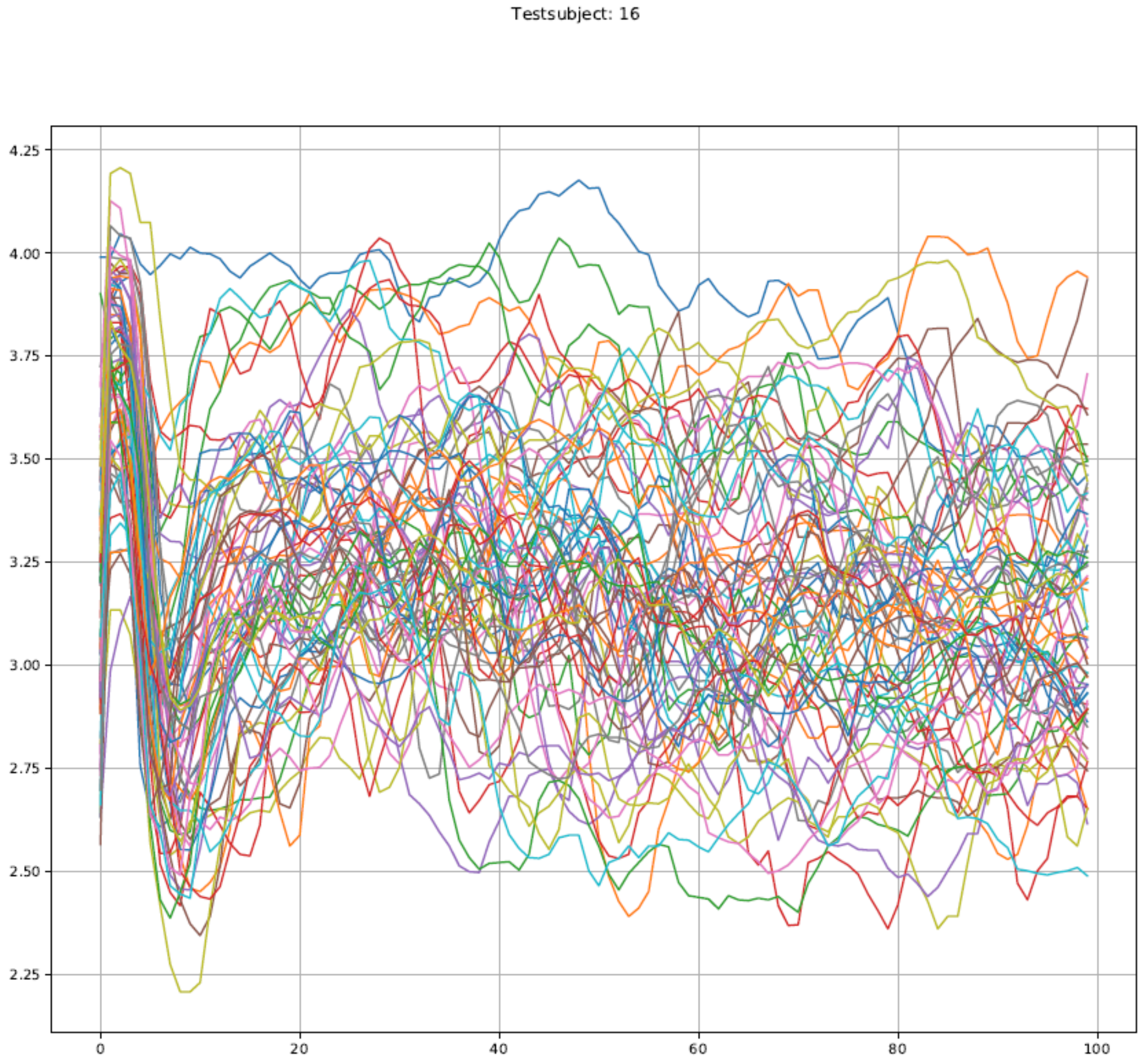


Figure 6.1: The overlapping plots of 70 PD signals of one test subject

These resting periods are meant to give some resting time for the test subject and also for the purpose of bringing the pupil diameter to its neutral state where the illuminance from the screen is low and does not force a decrease of the pupil diameter. The moment an IAPS picture is shown, the sudden increase of the illumination from the screen causes the pupil to constrict, causing the recorded signal to rapidly descend before bouncing back after the pupil starts to the illumination level of that picture. One factor that plays a significant role here to cause the pupil to dilate is the arousal reaction from seeing the stimulus picture. The more stimulating the picture is, causing more arousal, the faster the pupil dilates. This notable feature appears as well in a similar study [BMEL08].

To make it more systematic and easier for us to refer to each part of the PD signals components, we have proposed a custom nomenclature that will be used throughout this work.

6.1.1 PQR features

We have separated the typical pupil diameter behavior into three separate sections and designated three important landmarks which we have labeled as points P, Q, and R. P is the point where the peak is located followed by point Q where the pupil starts to dilate and then comes point R which is located 10 samples after point Q. The advantage of separating this behavior into three sections is that we can then observe how each section is affected by different levels of arousal. To further characterize the response, we extract three more parameters defined on the basis of the P, Q, and R points. These parameters are ΔPQ, m_{QR}, and ΔQR, and they are calculated as shown in Equation 6.1 where ΔPQ is the vertical difference between point P and Q, m_{QR} is the slope from point Q to R, and ΔPQ is the vertical difference between point Q and R.

$$\Delta PQ = P_y - Q_y \tag{6.1a}$$

$$m_{QR} = \frac{R_y - Q_y}{t} \tag{6.1b}$$

$$\Delta QR = R_y - Q_y \tag{6.1c}$$

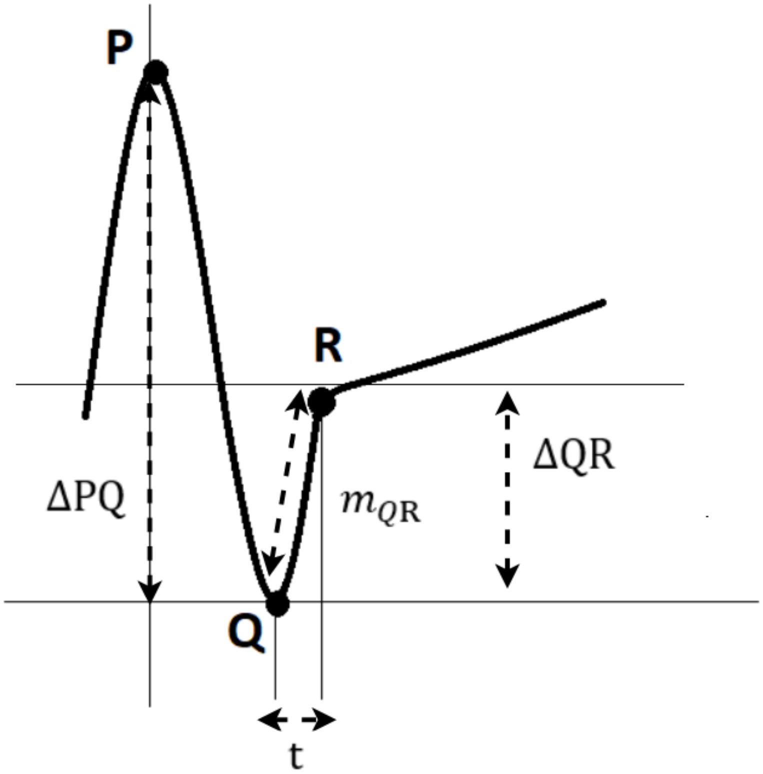
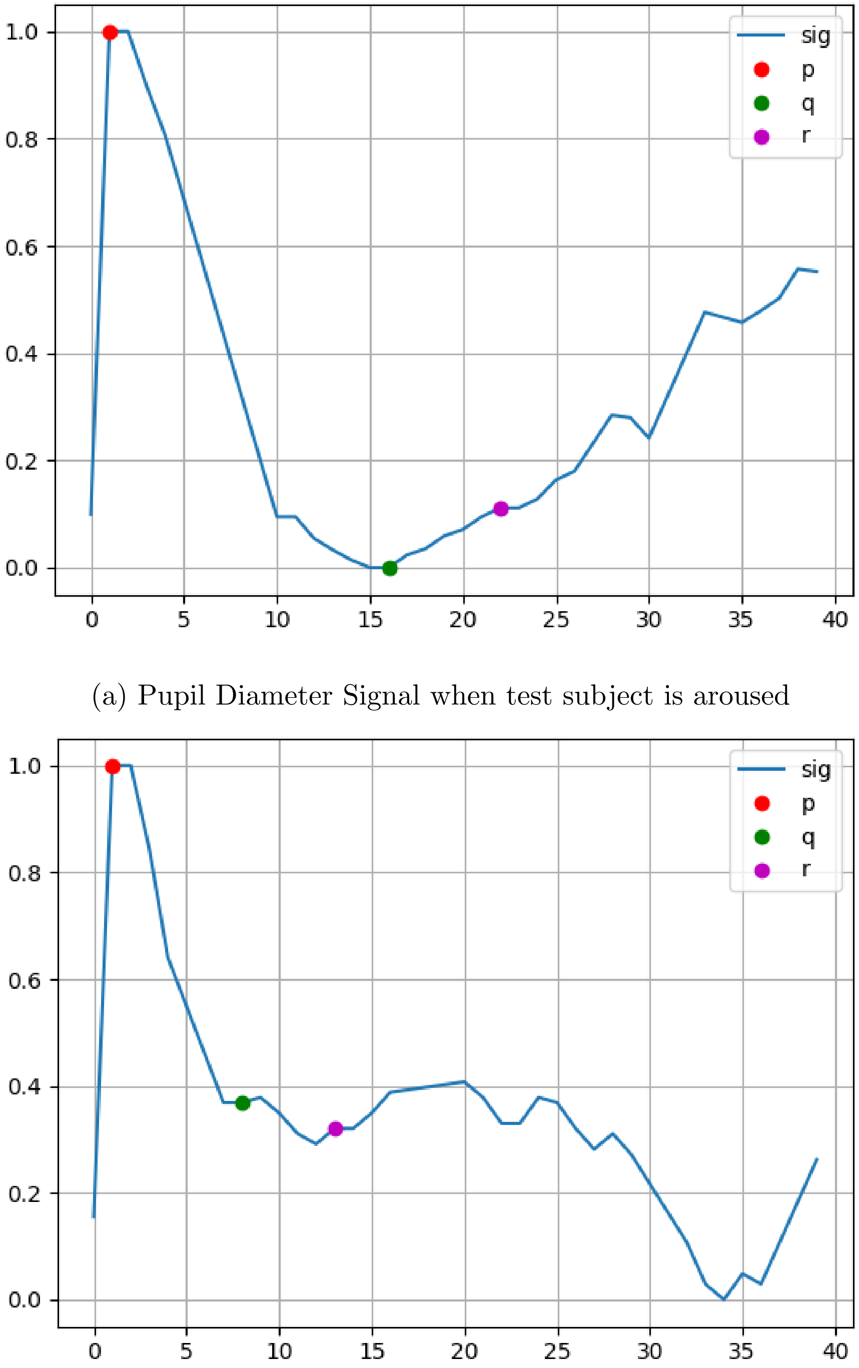


Figure 6.2: Diagram showing components of Pupil Diameter signal with denotations

6.1.2 Influence of arousal on PQR feature

Reviewing the data collected, the feature that is influenced by an arousal level the most is m_{QR} where the slope is larger in positive value when arousal is high. Otherwise, the slope is near zero or may even become a negative value. This is



(a) Pupil Diameter Signal when test subject is aroused

(b) Pupil Diameter Signal when test subject is not aroused

Figure 6.3: Demonstration of PD signals with each of them are marked the position of point P, Q, and R

reasonable given that the more stimulated a test subject becomes, the faster the speed of the pupil dilation; which causes the m_{QR} to have a sharp upturn behavior. While, in the low arousal case, the m_{QR} tends to have a slope value near zero. This distinct behavior of these two parameters makes them suitable to use as features for the predictive model.

6.1.3 Preprocessing Pupil Diameter Signal

In spite of efforts made during the experimental sessions to minimize the insertion of noise and artifacts in the data collected, these unwanted signal components still appear in the recorded signals. They may be caused by the test subjects movement, glitches from the device, or eye blinking. That is why we need to pass the data through a preprocessing pipeline to obtain data that is still usable.

In this section, we will describe the process of preprocessing of the Pupil Diameter signal including the detection of artifact and invalid data points and the explanation of the criteria to discard corrupted samples. We will also summarize how we extract PQR features to serve as inputs to the predictive model.

Merging Pupil Diameter signals from left and right eyes

The preprocessing the pupil diameter signal (PD) starts by discarding invalid data points. It is well established that, generally, the pupil diameters from both eyes are roughly equal and follow the same trends when they are constricted or dilated, influenced by many possible factors [Duc]. Often times, the eye-gaze tracker device losts track of one of the PD signal from either of the eyes. So, it is a reasonable approach to merge the left PD signal and the right PD signal together to obtain a more stable signal. The side that has the higher value is selected as part of the merged PD signal as shown in Equation 6.2

$$PD_{merge}(n) = max(PD_{left}(n), PD_{right}(n)) \tag{6.2}$$

Cropping the Interval of Interest

One record of the pupil diameter signal was obtained from the scenario where a test subject is presented with a picture from IAPS and then proceeds to rate its valence and arousal in response to the picture. All these actions happened within the total picture presentation time interval of 10 seconds (followed by 5 seconds rest with a gray screen) before the next IAPS picture is shown, as indicated Figure 6.4. Data are collected every 0.1 seconds so, in each record of a picture presentation(10 seconds) , 100 values of a PD signal are recorded. During this 1second interval the test subjects were asked to stand still as much as possible for the first 5 seconds and then they were asked to rate SAM rating in the remaining 5 seconds. In consequence, the last half of the record is when the test subjects were focused on rating their affective responses, looking and clicking on the SAM tool. Because of this, the analysis will not be performed in the last half of the 10-second interval. Instead, the analysis is performed on a cropped interval which includes from the 0th to the 39th samples which can be roughly estimated as the first 4 seconds after the IAPS picture is shown. We have found that our selected range contains the information that we seek and this choice is consistent with other related studies [BMEL08].

Artifact Identification and Linear Interpolation

Eye blink artifacts occur frequently in the recording of PD signals, when the Eye-Gaze tracker loses track of the pupil diameter while the eyelids down. Mariska

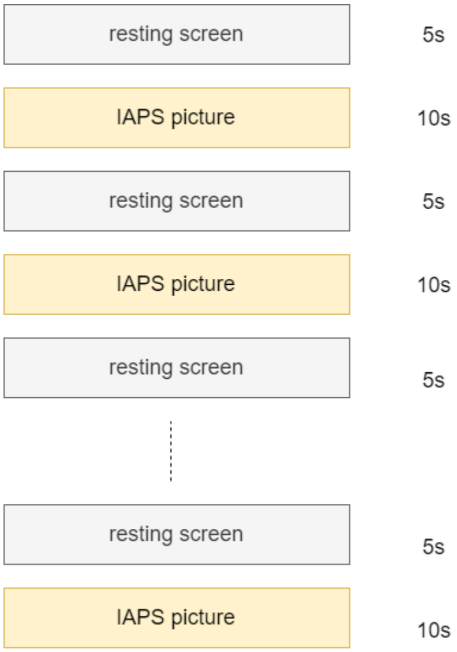


Figure 6.4: Scenario on how samples are presented to test subjects

[KSS18] has proposed a good practice on how to preprocess the pupil diameter signal which we have adopted for the identification of the artifacts. The common nature of artifacts is the abrupt change of the absolute pupil diameter size that is out of proportion to the adjacent data points. We call data points that exhibit this kind of behavior Dilation speed outliers. To define which portions of the record may be outliers, a normal dilation speed is needed to be postulated, in order to determine when disproportionate changes occur, implying an outlier. One cannot assume that the dilation speed throughout the complete record will be constant. Therefore, a normalized dilation speed, which is the maximum dilation speed at each sample relative to its preceding or succeeding samples, is used instead. The dilation speed can be calculated using Equation 6.3 where $d'^{[i]}$ is the dilation speed at each sample and $d[i]$ is the pupil diameter corresponding to the time stamp $t[i]$.

$$d'^{[i]} = max\left(\left|\frac{d[i]-d[i-1]}{t[i]-t[i-1]}\right|, \left|\frac{d[i+1]-d[i]}{t[i+1]-t[i]}\right|\right) \tag{6.3}$$

Then we proceed to calculate the Median Absolute Deviation (MAD) which will be used as a metric to detect the dilation speed outliers. MAD and the corresponding threshold value can be calculated using Equation 6.4 and Equation 6.5

$$MAD = median\left(|d'[i] - median(d')|\right) \tag{6.4}$$

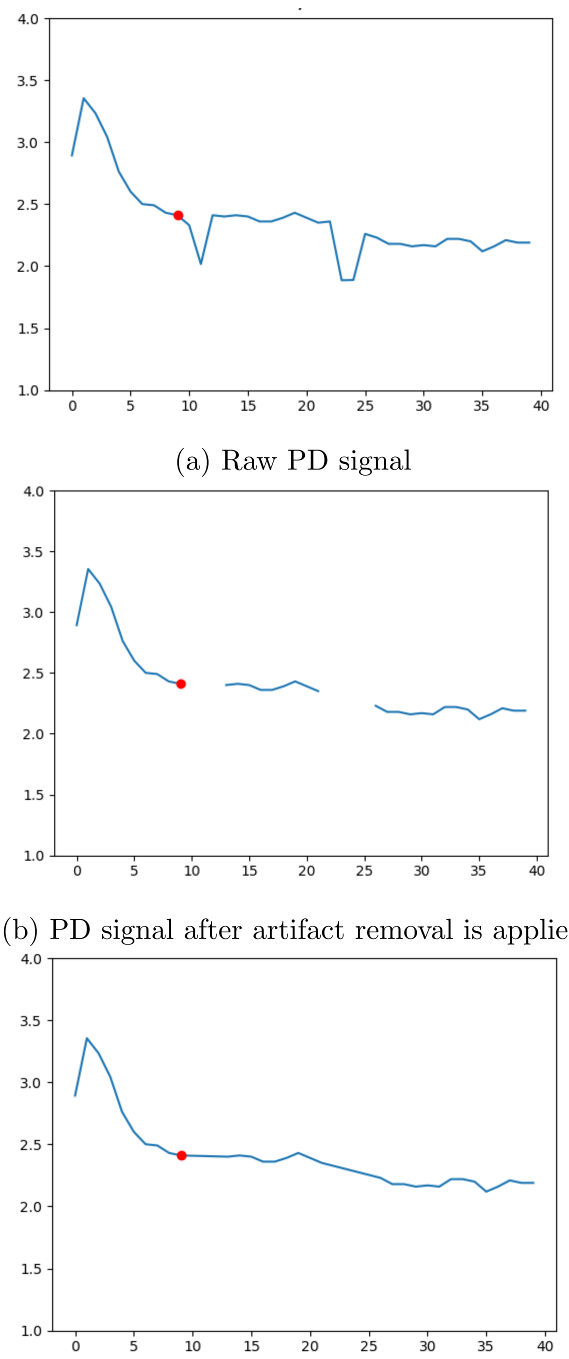
$$Threshold = median(d') + (n \cdot MAD) \tag{6.5}$$

The constant n is a control parameter that needs to be adjusted according to each application. Samples that have a dilation speed that is higher than the threshold will be considered outliers and should be discarded from the rest of the samples.

After the removal of the invalid samples, the recorded PD signals are now left with empty spaces that need to be padded to make use of the rest of the record, which still holds valuable information. However, records that have more than 25% of invalid data will be discarded. For the padding process, we choose to use the linear interpolation [Bre12] between the last sample before the invalid range occurs and the first valid sample after the invalid range. Figure 6.5 illustrates how each of these steps is applied to the PD signals.

Scale Normalization

Before we identify the location of the landmark points P, Q, and R, we have normalized the scale of the signals using the Minmax normalization method where the minimum value and the maximum value are obtained from the same test subject. This way, we can compare the changes in behavior at the normalized scale and eliminate the effect of pupil size changes due to subtle movements from a test subject. Equation 6.6 is the normalization equation used.



(a) Raw PD signal

(b) PD signal after artifact removal is applied

(c) PD signal after padded with Linear interpolation method

Figure 6.5: Demonstration of PD signal in each step of the preprocess method. Figure 6.5a shows raw PD signal with the glitch. Figure 6.5b shows PD signals after the MAD filter is applied and the glitches are removed. Fig. 6.5c shows PD signal after getting padded by Linear Interpolation method.

$$PD_{norm}(n) = \frac{PD(n) - min(PD_{per\_sbj}(n))}{max(PD_{per\_sbj}(n)) - min(PD_{per\_sbj}(n))} \tag{6.6}$$

Extracting PQR Features

After we obtained the preprocessed PD signal, each frame of PD signals resulting from the viewing of an IAPS picture is analyzed to identify the points P, Q, and R. As we described earlier, P is located at the peak of the overshoot so we simply identify point P at the maximum value in the first one-second interval (10 PD values) from the beginning of the data frame. Next, we can locate point Q by observing the first sequence after point P that changes its slope direction from descending to ascending. We achieve that by first applying the differentiation to the PD signal to obtain the trend of the slope. Subsequently, we apply the threshold at zero and convert the resulting differentiated PD signal into a binary signal where 0 indicates a downward turn and 1 indicates an upward turn. Then we just grasp the first sample index that turns to 1 after the location of P point and designates that index as the Q point. R is defined by simply locating the tenth sample after point Q, just to observe how the pupil changes after the initial onset of the record pass. Figure 6.6 summarizes the process of identification of the points that define this PQR complex. Also Figure 6.3a and Figure 6.3b shows the resulting plots of this algorithm.

The next necessary task is to obtain the features described in Section 6.1.1. Applying Equation 6.1 to PQR complex found, will yield the values of the desired features (ΔPQ, m_{QR}, ΔQR).

The diagram showing the whole preprocessing process for the Pupil Diameter signal is shown in Figure 6.9 which displays a birds-eye view of the process, as it was just described.

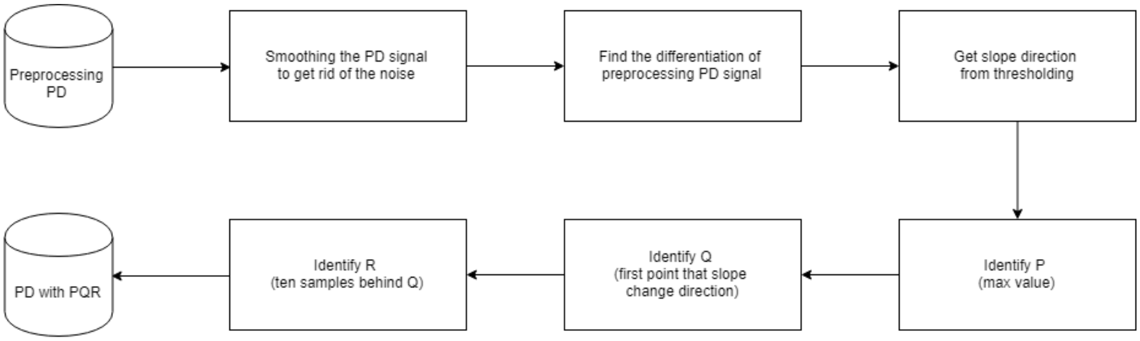

Preprocessing PD
Smoothing the PD signal to get rid of the noise
Find the differentiation of preprocessing PD signal
Get slope direction from thresholding
Identify P (max value)
Identify Q (first point that slope change direction)
Identify R (ten samples behind Q)
PD with PQR

Figure 6.6: Diagram showing the process of identifying PQR complex

Illumination Compensation

Previous chapters have already explained the need to account for the effect of Illumination on the pupillary response. We have discussed extensively in Chapters 3 and 4 a previously proposed approach to address this issue. However, when that previous approach was attempted on PD data contained in the short recordings that correspond to each IAPS picture presentation, it was found that an alternative strategy was necessary. Instead of using the Adaptive Noise Canceller to compensate for illumination changes in the PD signal directly (as a whole), we now were more interested in how the illumination affects the PQR feature components. In the similar study, led by Bradley [BMEL08], the authors present a comparison of PD response signals obtained when IAPS pictures with different levels of illuminance were presented to subjects, as shown in Figure 6.7. We notice that this figure is consistent with the characterization of the PD response as proposed at the beginning of this chapter: Higher luminance causes the initial PD interval to drop lower than moderate and low luminance. This minimum in the PD signal is the point we have identified as Q in the PQR complex. Conversely, point Q can be expected to be in a higher position when the IAPS picture we present to test subject has lower luminance than others.

Therefore, the position of the Q point is very important, because it affects the value of all the features extracted from the PQR complex. For example, if point Q is shifted down in the y-axis due to high luminance scenario, m_{QR} will get a sharper slope. Therefore, we propose that, by compensating the position of Q we can enhance the detectability of the high arousal effect presented in m_{QR} in low luminance cases, while reducing the possibility of mistaking the effect caused by high luminance with the effect caused by arousal in the case of a high luminance case. Consequently, based on the value of measured luminance for each IAPS picture sample, we compensate the location of point Q using Equation 6.7, where Q'_y is the adjusted vertical level of the point Q, Q_y is the original vertical level of point Q (before adjusting), α is the adjusting scale for fine tuning, $Illum_{max}$ is the maximum average luminance of all IAPS pictures used in the experiment, and $Illum_{sample}$ is the average luminance on the particular IAPS picture presented during the recording of the PD signal being considered.

$$Q'_y = Q_y - \left(\frac{\Delta QR}{\alpha} \times \frac{Illum_{max}}{Illum_{sample}} \right) \tag{6.7}$$

The idea here is to shift the position of Q up and down based on the average luminance measured from each of the IAPS pictures. If the luminance of a particular IAPS picture is low in comparison to the value of the maximum luminance of all IAPS pictures used in the experiment, then the position of point Q in the y-axis will be shifted down. Equation 6.7 is designed in a way that point Q will always shift down when compensated; however, the compensated value will be different based on the average luminance of each IAPS picture sample. The parameter α is for controlling the range of compensation. For instance, if α is set to 2 then the maximum compensated value (the range that point Q will be shifted down) will be half of ΔQR. By tuning hyperparameter, α, one can control the effect of illumination compensation on the dataset. Subsequently, after the position of point Q in the y-axis is adjusted, all PQR feature values are re-calculated. Figure 6.8 shows the sup-process followed for illumination compensation.

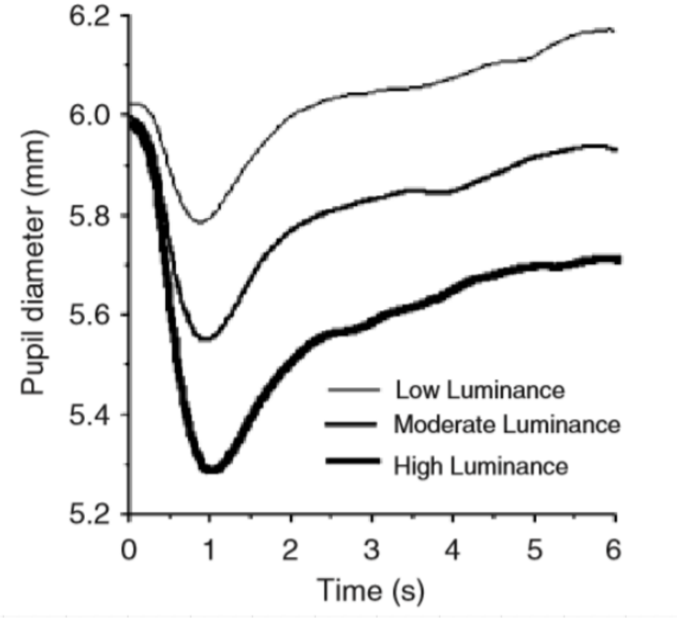


Figure 6.7: Comparison of PD signals affected by different level of Illumination. Picture from [BMEL08]

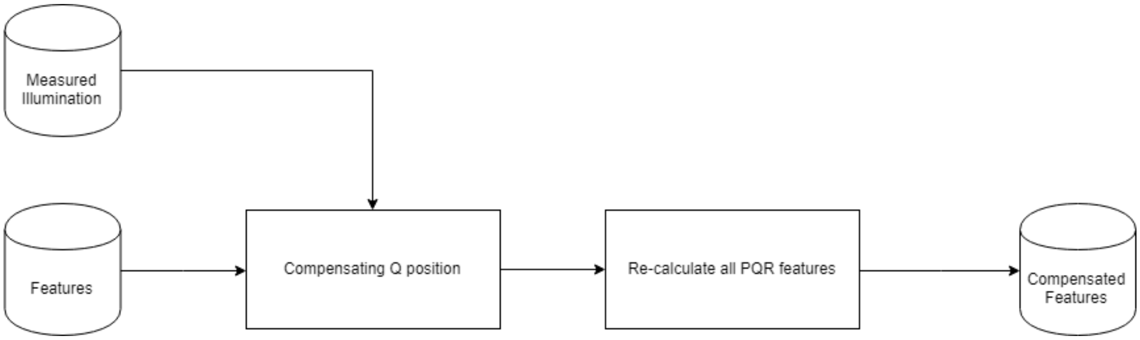


Figure 6.8: Diagram showing process of Illumination compensation

Figure 6.9 demonstrate the pipeline of the whole process of preprocessing of the pupil diameter signal. Our goal as to obtain PQR features (ΔPQ, m_{QR}, and ΔQR) is achieved at this point.

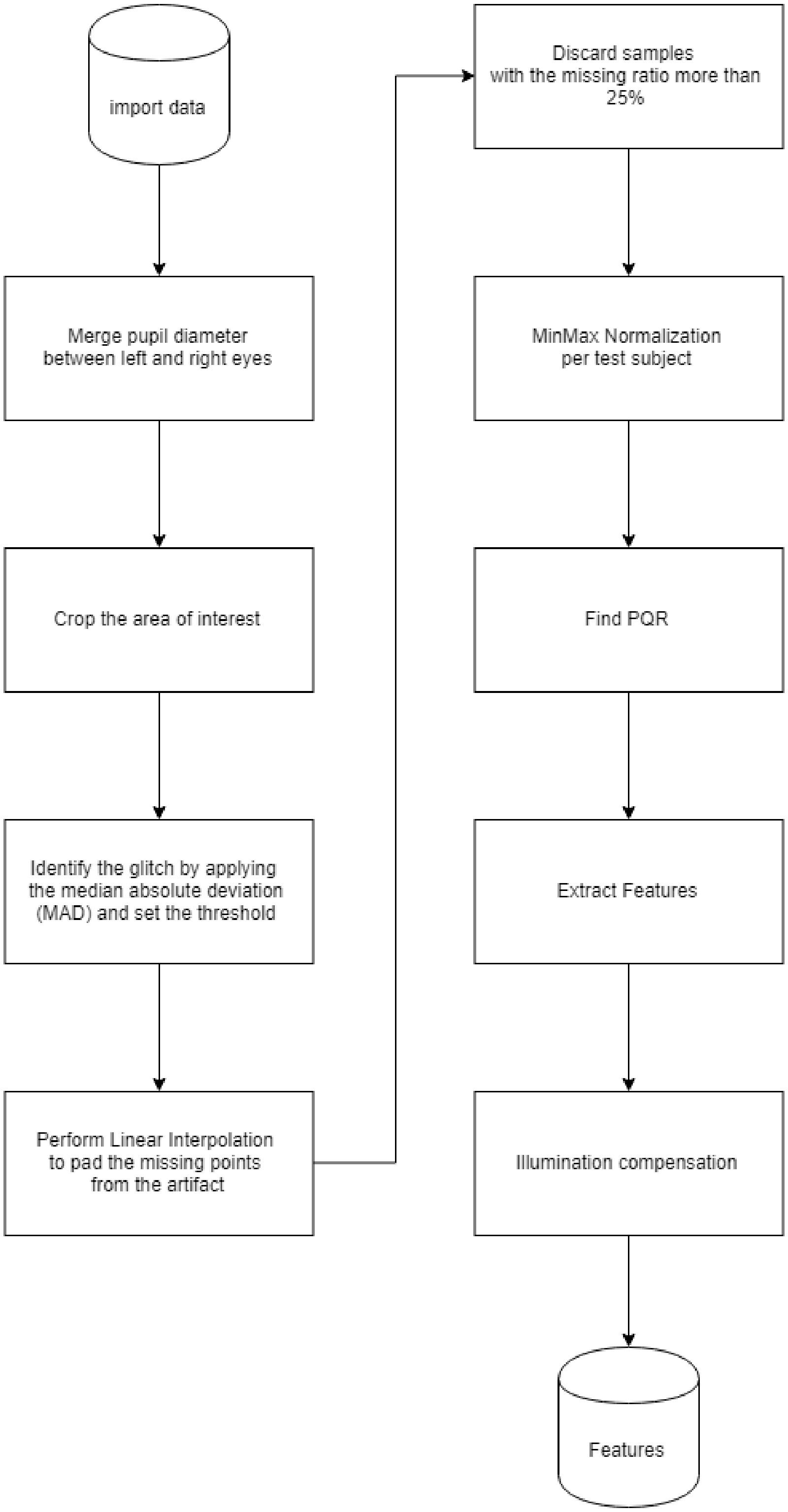

import data
Merge pupil diameter between left and right eyes
Crop the area of interest
Identify the glitch by applying the median absolute deviation (MAD) and set the threshold
Perform Linear Interpolation to pad the missing points from the artifact
Discard samples with the missing ratio more than 25%
MinMax Normalization per test subject
Find PQR
Extract Features
Illumination compensation
Features

Figure 6.9: Diagram showing the pipeline of Pupil Diameter signal preprocessing

6.2 Facial Action Parameter Vector (FAP Vector)

In this section, we turn our attention to the type of data that we collected to estimate the valence parameter. First, we will discuss the nature of the FAP vector signal and discuss the kind of features we want to extract from it.

In Chapter 3 we have already explained in detail the definition of each FAP measurement. All the FAP measurements are listed in Table 3.3. As a summary, we recall that each FAP vector is the measurement of the difference between two points of a group of designated facial points which we consider as important landmarks to encode the facial expression information in an efficient way that also has an advantage in terms of the memory management. When a unique facial expression occurs, for example, a smile, those facial points are moved based on the contraction of specific facial muscles and the effect of that transition is reflected in each FAP vector. Here, we are focusing on the behavior of FAP vectors regarding the movement of facial muscles and we categorize those responses into three categories.

- Shrink (decrease in value, denoted as -1)
- No movement (no change, denoted as 0)
- Extend (increase in value, denoted as 1)

Shrink means that two facial points are getting closer to each other while Extend implies the opposite, and, for the case when there are no distinct changes we will conclude that there is no movement, in that FAP vector. By systematizing the observations it this way, we can more easily study the modifications in FAP vectors. Furthermore, we can map groups of FAP vector responses to each action unit (AU) (described in Chapter 2) in order to consider which action units are activated during the emergence of a specific facial expression. Consider Table 6.1, which shows the relationship between FAP indices and AUs. An example that illustrates the use of this table to identify the activation of AUs is found in the first row of Table 6.1. This row contains AU1, which is a unit describing the raising of the inner brow. To regard AU1 as activated, both FAP indexes 31 and 32 have to be in the extending state; otherwise, AU1 is not activated. The entries in the Signal Index column are simply labels that are used to identify the different FAP traces in multi-trace plots.

The purpose of the encoding presented above is to facilitate the mapping of facial expressions to the valence scale. There is not a direct connection (like the one between pupil diameter and arousal). Therefore, investigate associations between the kinds of facial expression and the valence scale. In Chapter 2, we have already introduced Russells Circumplex Model of Affect (Figure 2.1) which includes the placement of some basic emotions on its circular graphical representation. This provides us with a rough idea of where those facial expressions should be on the valence scale. In Chapter 2, we introduced the relationship between Action Units (AUs) and facial expression (refer to Table 2.1) and because we already have a way to convert our FAP vectors to AUs, we now have a way to imply the position of a facial expression in the valence scale, based on the behavior of FAP vectors.

That is why the features that we will extract from FAP vectors will be the activation response of AUs based on the behavior of the FAP vectors. Next, we will describe how we arrive from the raw FAP vectors to the AUs activation response.

6.2.1 Preprocessing Facial Action Parameter Signal

Just as the PD signals required some pre-processing manipulations, the facial expression data generated by Kinect also requires pre-processing.

Table 6.1: Relationship between Action Unit (AU), Facial Action Parameter (FAP) and Signal Index

AU	AU_description	Signal_Index	FAP_Index	Shrink _Extend
AU1	Inner brow raiser	0	31	1
		1	32	1
AU2	Outbrow raiser	2	35	1
		3	36	1
AU4	Brow lower	0	31	-1
		1	32	-1
		4	37	-1
		5	38	-1
AU5	Upper lid raiser	6	19	1
		7	20	1
AU6	Cheek raiser	6	19	-1
		7	20	-1
		8	41	-1
		9	42	-1
AU9	Nose wrinkler	10	61	-1
		11	62	-1
AU10	Upper lip raiser	12	59	-1
		13	60	-1
AU12	Lip corner puller	12	59	-1
		13	60	-1
		14	53	1
		15	54	1
AU15	Lip corner depressor	12	59	1
		13	60	1
AU16	Lower lip depressor	16	5	-1
		18	16	-1
AU20	Lip stretcher	14	53	-1
		15	54	1
		16	5	-1
AU23	Lip tighter	14	53	-1
		15	54	-1
AU26	Jaw drop	18	3	1
		16	5	1

Conversion form FAP vectors to FAP units

As we mentioned in Chapter 3, we have measured the FAPU parameters on each test subject before we started the experiment session. The purpose of this unit is to eliminate the distortions caused by the inconsistent facial size among test subjects. The raw FAP signals then are divided by FAP units as described in Table 3.2.

Cropping the Interval of Interest

Similar to the PD signal case, the interval of interest for feature extraction from the facial expression data is only the first 60 sequences (6 seconds) of the record. We have isolated the signals only from that interval, for analysis.

Eliminating Impulsive noise

Examples of the raw facial expression data (evolution of the 19 FAPs through the presentation of an IAPS picture) can be observed in Figure 6.10. It can be noticed that the signals are severely corrupted by impulsive noise. Accordingly, we decided to apply a Savitzky-Golay filter which is a digital filter that is well- known of its smoothing property [S$^+$11]. The filter works by connecting two adjacent samples together with a polynomial curve. This helps smooth the signal without significant distortion on the original trend of the signal. Fortunately, there is an available toolbox for the Savitzky-Golay filter in the Scipy package [Bre12]. We have applied this smoothing filter to reduce the impulsive noise in the FAP data.

Discarding Invalid Samples

In some instances, Kinect lost track of the face of a test subject resulting in the recording a constant FAP value during that face-loss interval. We have regarded these occurrences as missing part of the sample. We have discarded the samples that have more than 25% of the FAP data missing.

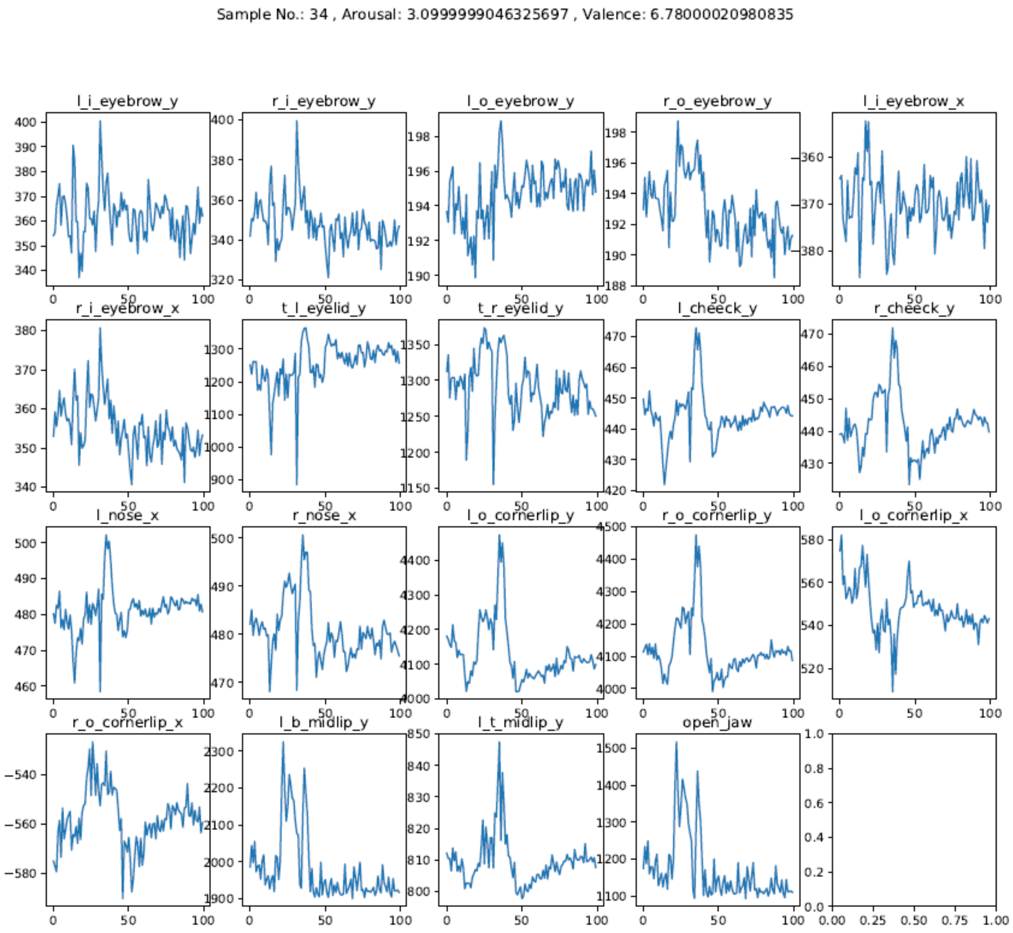


Figure 6.10: Plots of each raw FAP vector signal

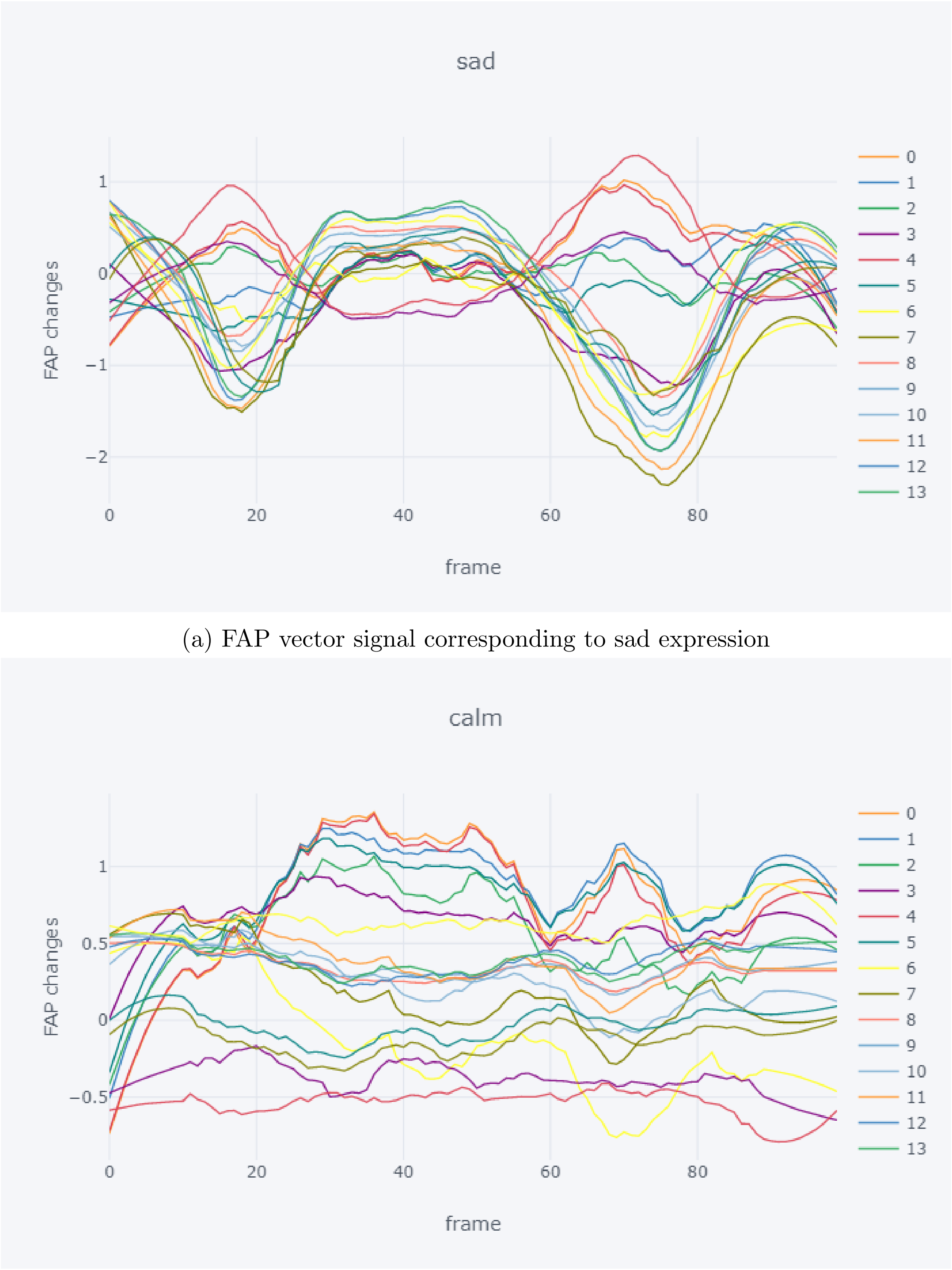
Scale Normalization

Each of the FAP vectors is measured from different parts of the face. As different subjects may have faces that vary in the size of some of their dimensions (e.g., broad faces vs. narrow faces) the FAP values need to be normalized. Feeding features without standard normalization to a machine learning process might create some biases in the model. To avoid this issue, we have performed the standard normalization, where the mean offset of the FAP vectors is removed they are scaled to unit variance. This is a useful way for a statistician to compare values from different units on the same scale, as suggested in [PVG$^+$11].

After the pre-processing steps mentioned, the feature extraction process can be applied to the FAP vector signals. The description of the feature extraction process in the following sections will make reference to observations made on plots of FAP vector signals on the same graph identified by signal index (refer to Table 6.1) in the plot legend. In order for the reader to see the different response from FAP vectors corresponding to several types of facial expression, we have included the plots of FAP vector signals for a variety of types of facial expressions, such as surprise, laugh, disgust, anger, fear, etc. All these plots can be found in Figures 6.11, 6.12, and 6.13.

Identifying the Temporal Occurrence of Facial Expressions

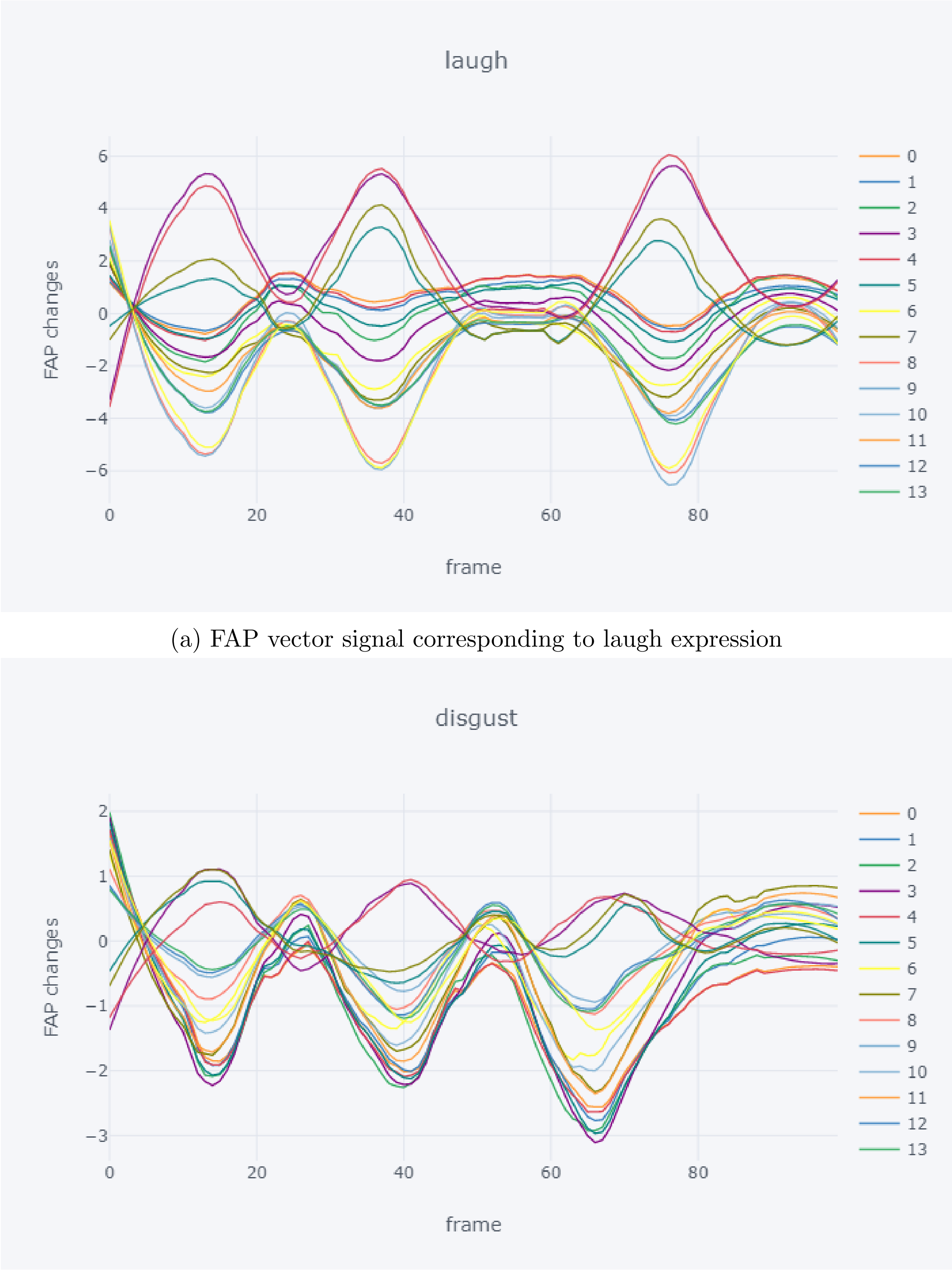
The behavior that we found from studying the data collected is that a facial expression evolves in time through three states: onset, peak, and retreat. Most of the time the expression lasts between 2 to 4 seconds and its occurrence is marked by visible correlations between FAP vectors. The simplest behavior that we use for the first detection of an occurrence of facial expressions is when several of the FAP vectors change with the same, simultaneous trend. To detect these instances we used the peak finding python library from [Neg18] However, several preliminary steps were performed before we applied the peak finding algorithm. First, the absolute value of all the FAP vector signals is taken, since the algorithm can only detect positive peaks. Second, we sum all FAP vector signals together to amplify the size of coordinated changes. Then, we apply the peak finding algorithm. This algorithm uses an attribute named threshold which is the parameter used to impose a normalized threshold (ranges between 0 to 1). The peaks that are not higher than the threshold will not be counted as peaks. For example, if the threshold attribute is set to 0.5, peaks that have a value in the normalized scale below 0.5 will be discarded. We have



(a) FAP vector signal corresponding to sad expression

(b) FAP vector signal corresponding to no expression

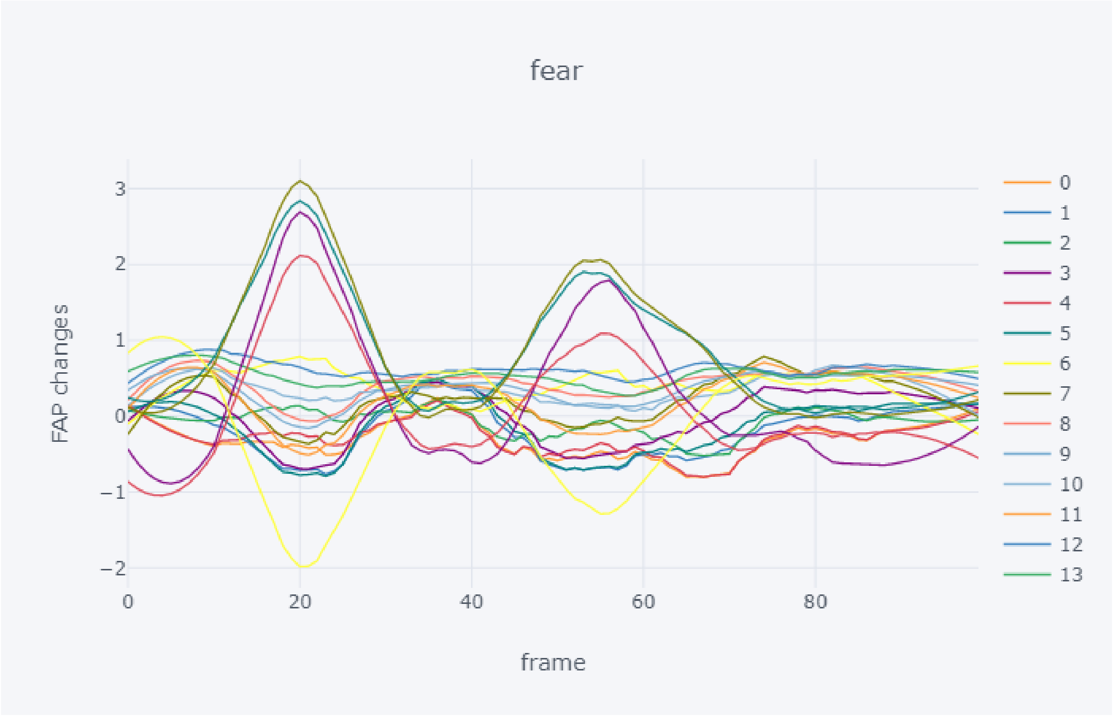
Figure 6.11: Plots of FAP vectors after removing noise and standard normalization



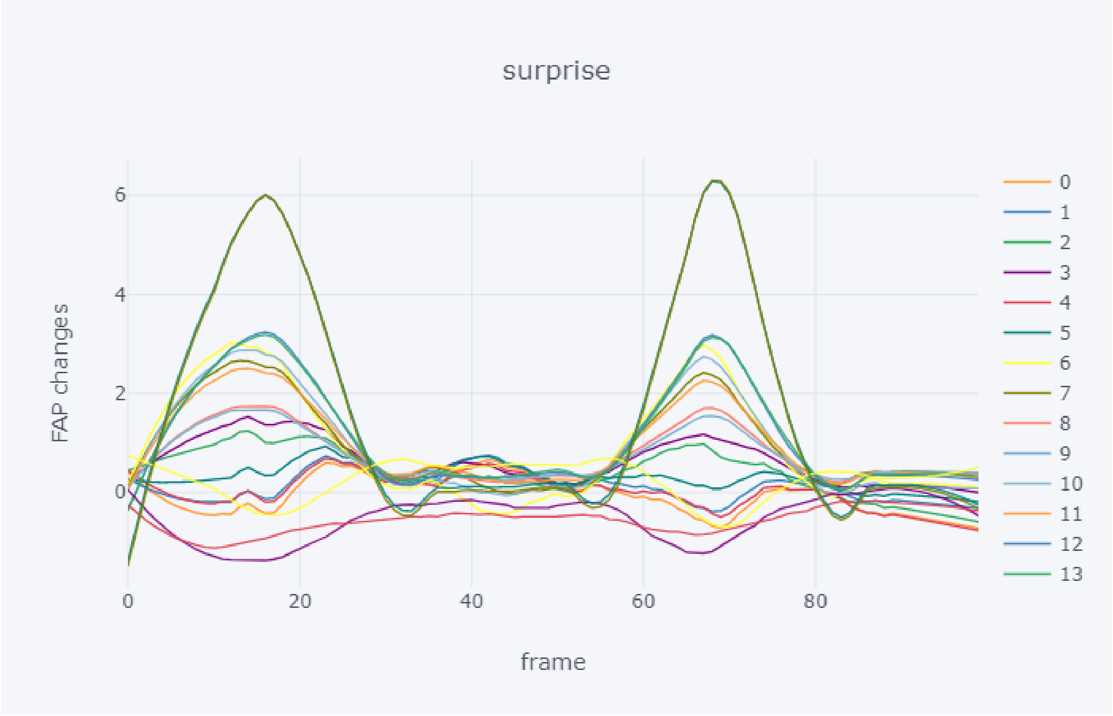
(a) FAP vector signal corresponding to laugh expression

(b) FAP vector signal corresponding to disgust expression

Figure 6.12: Plots of FAP vectors after removing noise and standard normalization

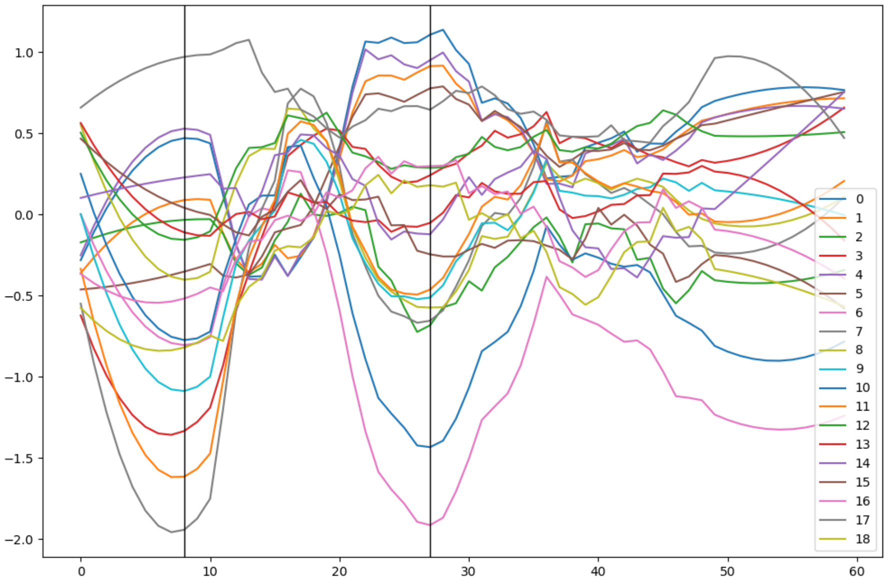


(a) FAP vector signal corresponding to fear expression

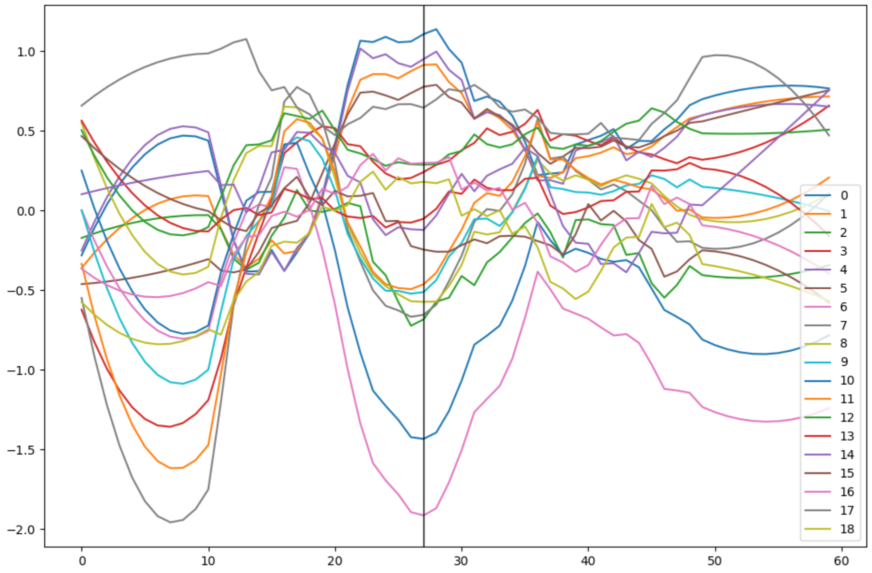


(b) FAP vector signal corresponding to surprise expression

Figure 6.13: Plots of FAP vectors after removing noise and standard normalization



(a) Potential facial expression moments



(b) Selected facial expression moment

Figure 6.14: Demonstration of facial expression moment selective process. Figure 6.14a shows the result of the first screen to find the potential facial expression moments. Figure 6.14b shows the selected facial expression moment corresponding to the applied criteria.

set this threshold to 0.6 in our case. In some cases, no peaks are detected. This scenario is interpreted as corresponding to a calmed (neutral) facial expression, which implies that there is no response from the test subject to the stimulus and that he/she did not change his/her facial expression significantly. It is very important to be able to identify when a calm" reaction occurs because this reaction implies the center position on the valence scale. The diagram of the processing sequence for facial expression location is shown in Figure 6.15. The results of these steps are lists of facial expression locations or indications that there were no facial expressions detected. An example of the results from the pre-processing sequence is shown in Figure 6.14a.

Select Facial Expression Moment

If more than one potential facial expressions are located within a single record the feature extraction process will focus on just the first reaction from the test subject after viewing the stimulus picture, as this is expected to be the spontaneous reaction to the stimulus. This is also why we only analyzed the signal during the first six seconds of each record. From the observation of numerous FAP vector plots it was estimated that signal deflections that represent legitimate face expressions usually have a base length in the range of 2 to 4 seconds. In contrast, signal deflections that exhibit shorter base lengths happen to be glitches produced by sudden movements of test subjects. Signal deflections with higher peak value are desirable as well, because they indicate an extensive movement of facial muscles. Combining those two desirable attributes together, we propose the Expression Detection Parameter (EDP) defined in Equation 6.8. When this parameter is high, it is likely that the corresponding peak is a facial expression in response to the stimulus. The base length and peak height attributes of the peaks found are provided by the PeakUtils

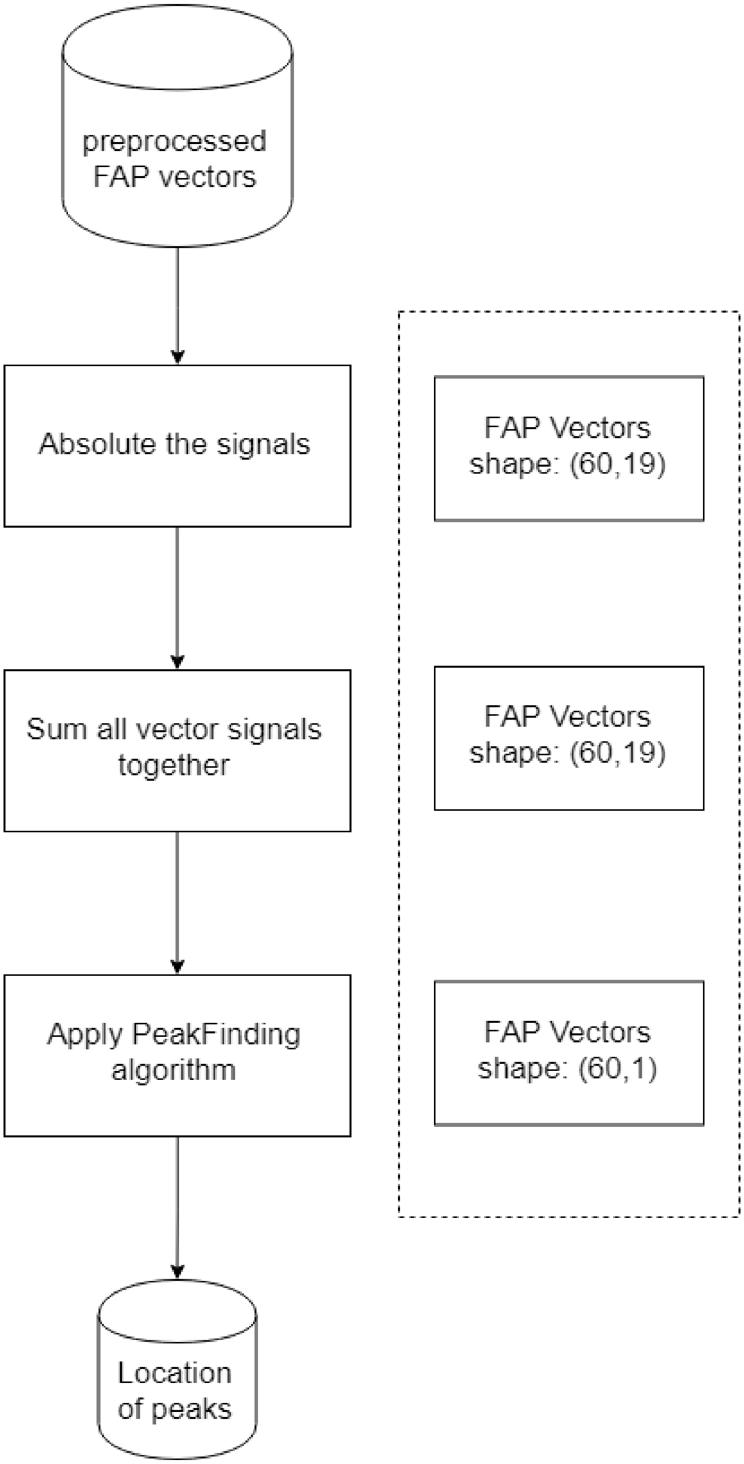

preprocessed FAP vectors
Absolute the signals
Sum all vector signals together
Apply PeakFinding algorithm
Location of peaks
FAP Vectors shape: (60,19)
FAP Vectors shape: (60,19)
FAP Vectors shape: (60,1)

Figure 6.15: Diagram showing process of identifying facial expression moments

library [Neg18], and the algorithm selects peak with the highest Expression Detection Parameter.

$$ExpressDetectionParameter = \frac{base\_length}{peak\_height} \tag{6.8}$$

Extracting Direction Vectors

In this step, we will extract the state of each of the FAP vectors, as stated earlier in Section 6.2. The peak-finding process described above was applied to the sum of all the FAP vector signals, just for the purpose of locating the key deflection of the signals in time. However, for the identification of which FAP signals exhibit significant deflections in the selected interval, the peak-finding algorithm must be applied to each of the FAP signals individually. This will identify the FAPs that truly became activated, and it will also show those FAPs that might have remained inactive. For example, FAP index 3 (Jaw drop) might not change in the surprise expression if a person does not open his/her mouth during the time when the facial expression occurs. By performing this additional step, we can identify the state of each FAP vector more accurately.

The direction vectors that are obtained next are identical to FAP vectors, except that they contain the state of FAP vectors at the peak of the facial expression. The value of direction vector can only be 1 (FAP vector extended), 0 (no change in the FAP vector), and -1 (FAP vector shrunk). This value is obtained by determining what type of peak occurs during a window of length 20, centered around the location of the facial expression moment, determined by the previous step. If the peak is a positive peak, we set the value to 1. If the peak is a negative peak, we set the value to -1. Finally, if no peak is detected, we set the value to 0. Figure 6.16 shows an example of the stem plot of the direction vector for the 19 FAP vectors. Once the direction vector is composed, the corresponding AU vector can be formulated.

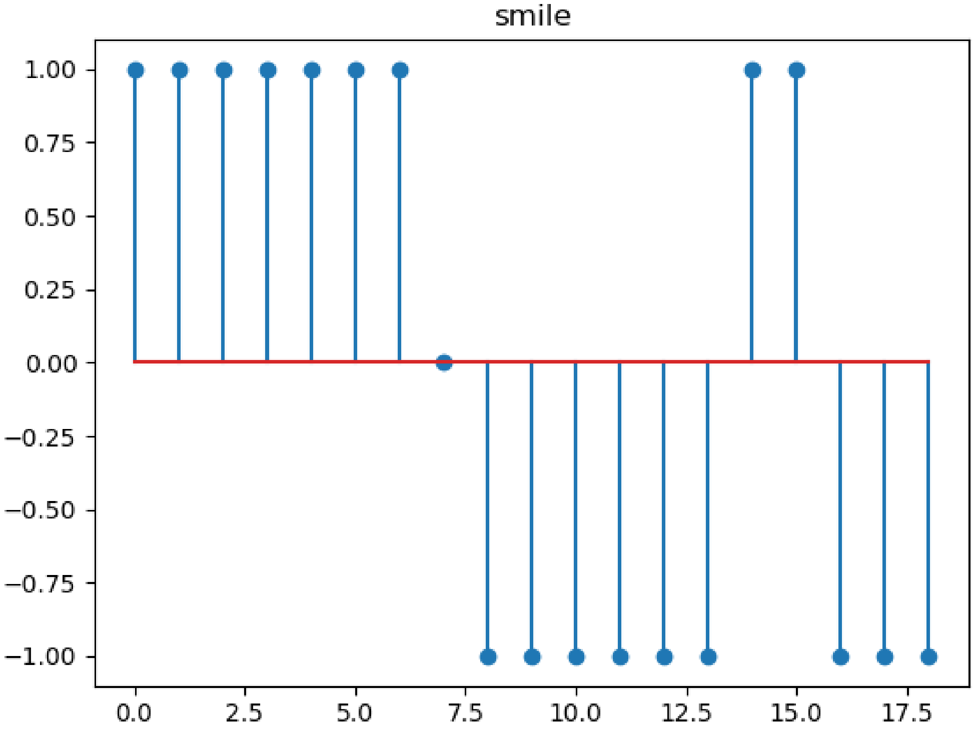


Figure 6.16: Stem plot of Direction Vectors

Generating AU vectors

AU vectors are generated according to the information contained in Table 6.1. As discussed in Section 6.2. AU vectors contain the state of Action Units (AUs) which can be 0 (not activated), or 1 (activated). Each of these vectors is conformed as 1 row and 13 columns, since we only observe the 13 AUs listed in Table 6.1. These AU vectors, derived from the corresponding FAP vectors, will be used as features for estimating valence trough a predictive model.

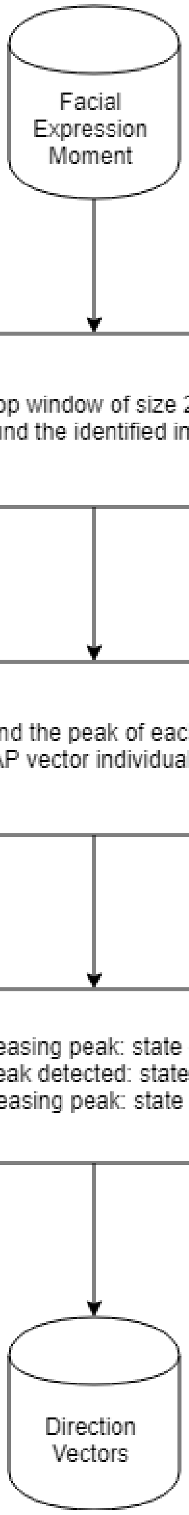

Facial Expression Moment
Crop window of size 20 around the identified index
Find the peak of each FAP vector individually
Increasing peak: state -> 1
no peak detected: state -> 0
Decreasing peak: state -> -1
Direction Vectors

Figure 6.17: Diagram showing the process of extracting Direction vectors

6.3 Summary

In this chapter, we have explained the manipulations used to pre-process the data from both sensors in the system, which including noise removal and normalization. We also discussed the observations made as the data were visualized and how these observations helped us in defining additional pre-processing steps. Furthermore, we discussed how we extract features which we think are important features that will enhance the predictive models performance.

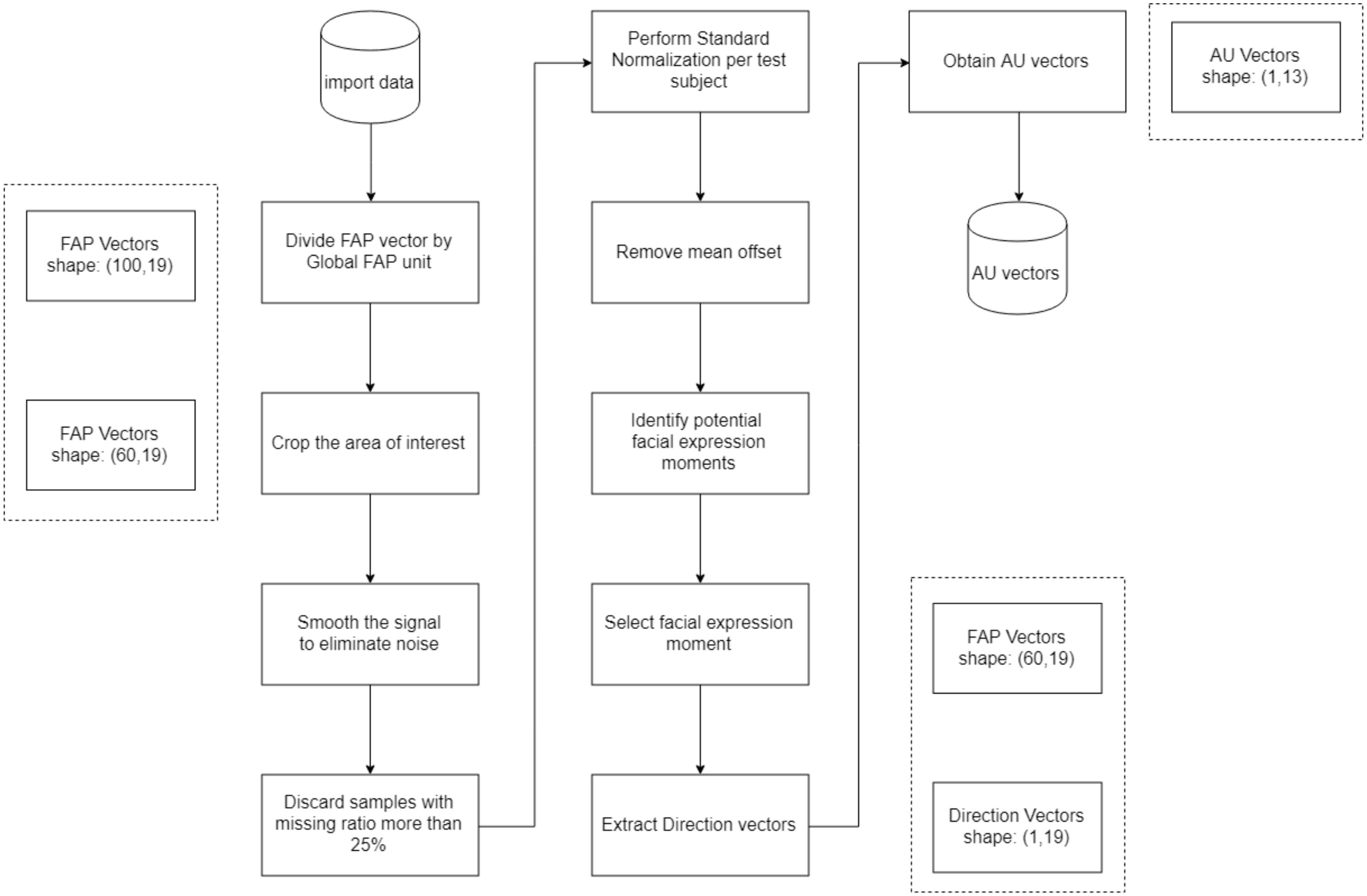

FAP Vectors
shape: (100,19)
FAP Vectors
shape: (60,19)
import data
Divide FAP vector by Global FAP unit
Crop the area of interest
Smooth the signal to eliminate noise
Discard samples with missing ratio more than 25%
Perform Standard Normalization per test subject
Remove mean offset
Identify potential facial expression moments
Select facial expression moment
Extract Direction vectors
Obtain AU vectors
AU vectors
AU Vectors
shape: (1,13)
FAP Vectors
shape: (60,19)
Direction Vectors
shape: (1,19)

Figure 6.18: Diagram showing the pipeline of Facial Action Parameter (FAP) Vector preprocessing. Blocks in dotted blocks show the shape, in a format (row,column), of the current state of the processed vector. (Shape:(100,19) means 19 FAP vectors with each vector has a length of 100 sequences.)

CHAPTER 7

PREDICTIVE MODEL

This chapter describes the approached followed to develop a predictive model that can estimate the arousal and valence experienced by a subject from the features described in Chapter 6. The data available for the development of the model we obtained from 50subjects who volunteered to participate in this study. Each participant was presented with 70 picture samples from the IAPS database and was asked to rate his/her arousal and valence levels according to the scales of the SAM tool. Thus, overall, there were 3,500 samples of pupil diameter and facial expression records (in the form of FAP vectors), and with the corresponding subjective SAM ratings of arousal and valence.

Figure 7.7 shows a bird-eye view of the whole process of building the predictive model for this study. Before the process followed for model building is explained, a brief overview of the machine learning approach selected, the Support Vector Machine (SVM), will be included. Then the specifications of the model used, as well as details pertaining to the assignment of features labels, will be discussed. Towards the end of the chapter, we will explain how we tuned the models hyperparameters to enhance its performance.

7.1 Model Architecture

In chapter 2, we have discussed the Circumplex Model of Affect introduced by Russell. In this model, the affective state of a subject can be estimated by obtaining approximate values of the corresponding arousal and valence. In pursuit of that goal, we propose to use tow cascaded classifiers. The first part of the model has the role of estimating the valence parameter from the FAP vectors before sending its estimated valence to the second part of the model which is the arousal-valence classifier. This second classifier takes features from the pupil diameter signals and the initial estimated valence as inputs to classify the affective state (valence-arousal levels) of the user. Figure 7.7 shows the diagram of the architecture of the predictive model used in this study.

There are two reasons behind the decision to design the model architecture in that particular way. First, separating the model into two parts will enable each part of the model to focus on the input features that actually matter as our hypothesis suggests: the facial expression as the indicator of the pleasure or displeasure (i.e., valence); and pupil diameter for identification of high or low arousal level. It is known that the pupil diameter signal cannot appropriately differentiate how happy or sad a user is. Therefore, it seems appropriate to not use it to drive the part of the system that is meant to identify the valence of the affective state (as it could act as a confounding factor, in that context).

However, it has been proposed that there is a level of correlation between the valence and the arousal of many affective states experienced by humans. As shown in Figure 3.4a, the 2-dimensional plot of the mean of arousal and valence ratings for each IAPS picture from the IAPS repository shows a very noticeable V-shape distribution across the plot. This suggests that high valence, both pleasure and displeasure, tends to appear with high arousal. However, prior knowledge of the arousal level would still leave the possibility of two likely levels of valence, in that V-shape. Thus, estimating valence first, and feeding the estimated valence to the second classifier to estimate arousal seemed to be an advantageous approach.

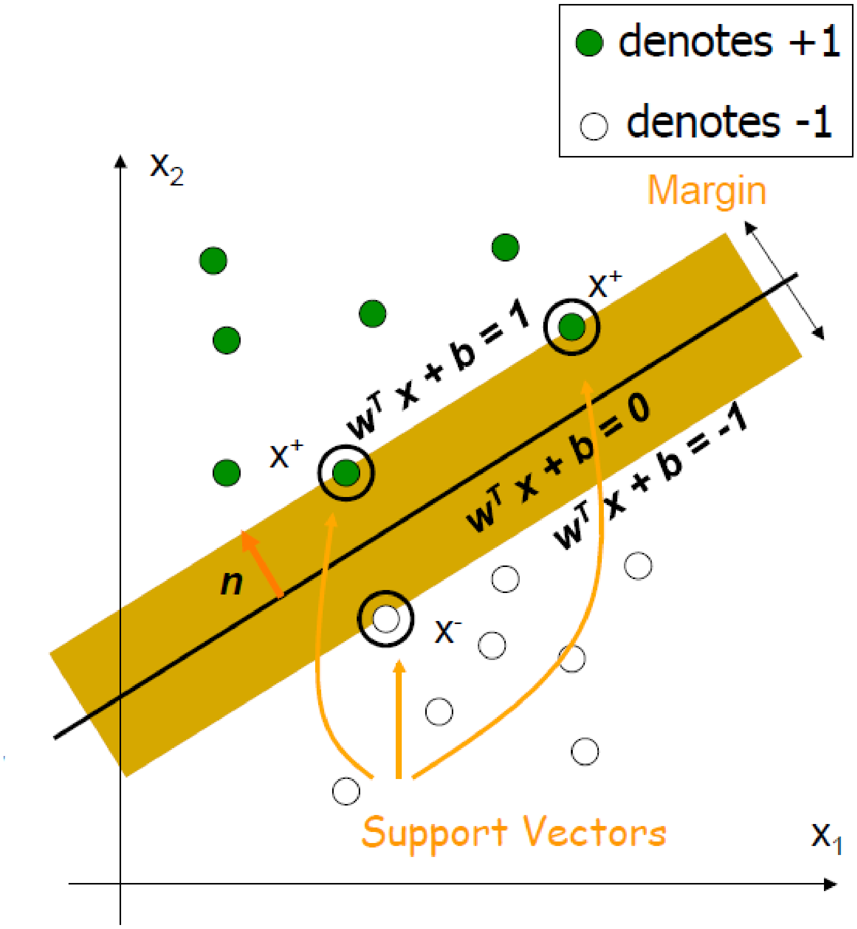


Figure 7.1: Illustration of Support Vector Machine and its element

7.2 Support Vector Machine (SVM)

In this study, we have chosen the Support Vector Machine (SVM) as the classifier model due to its simple, yet powerful method to classify the tabular type data. A brief review of Support Vector Machines is provided below.

Support Vector Machine (SVM) [MBD$^+$90] is a supervised machine learning model that is based on the idea of generating a separating hyperplane to discriminate the classes in the dataset. Data patterns belonging to different classes fall on the different side of the hyperplane. The hyperplane itself can be regarded as linear discrimination boundary in two-dimensions and as a hyperplane in higher dimensions. It is defined by the support vectors, which are data points that are closest to the hyperplane (a vector is another way to regard a data point). By controlling the regularization parameter (often referred to as C), which is the parameter to control the trade-off between margin and training error, we can control how much we want the model to avoid misclassifying each training example. If the value of C is set high, then the smaller-margin hyperplane will be a sought for the optimization of the model during training, and this can lead to the problem of overfitting. If overfitting is allowed, the model has a very high accuracy for the training data but achieves a lower accuracy for test cases not used during training. Another important characteristic of SVMs is their use of kernels, used to overcome SVM limitations in classifying data that are non-linearly separable in their native dimensionality. In those cases, a kernel, which is usually a mathematical function, such as Gaussian, Polynomial, Sigmoid, etc., is applied to map the data into a higher dimension, where it might now be possible to separate the classes with a hyperplane. SVM is still a popular machine learning approach, and it is used across many disciplines of research due to its robustness, high speed of training and simplicity of deployment.

7.3 Target Labels

The goal of the is study is to be able to assess the affective state of the user in the Circumplex Model of Affect. In order to achieve that goal, we have to estimate arousal and valence parameters (See Figure 2.5 for a visualization of this fact).

Because the scales in the Circumplex Model of Affect and in the SAM tool are different, we have to convert the SAM scale (which ranges from 1 to 9) to the range where the scale is compatible with the circumplex model, having its center is at the origin point. We then define the conversion equation as shown in Equation 7.1. The parameter A_scale is the scaling factor to control the range of the new scale which we set it to 1 at default. Applying this conversion, the maximum range of the SAM rating scale is mapped to the interval -4.5 to 4.5, with its center at the origin point.

Furthermore, to make this as simple as possible, we divide the area of the circumplex model into 4 quadrants and assign 4 labels to the quadrants, based on the value of arousal and valence, as shown in Figure 7.2.

The target label now will be converted again, one last time, to be either 0 or 1. If the target value is lower than zero then the target value is set to 0. In the other hand, if the target value is higher or equal to zero, it will be set to 1. Altogether, four labels which are HV (High Valence or 1), LV (Low Valence or 0), HA (High Arousal or 1), and LA (Low valence or 0) will serve as a target labels for the classification task.

$$Target\_Scaled = A\_scale \times \frac{(SAM\_Scale - 5)}{4} \tag{7.1}$$

7.4 Input Features

The two types of input features used to train the valence classifier and the arousal-valence classifier are presented here. These features will be explained visualized, to help understand them in a more intuitive way.

7.4.1 Features for the Classification of Valence

The main features that we use to feed into the valence-classifier are AU vectors (explained in Chapter 6). We discussed previously in Chapter 3 the guidelines used to select a subset of IAPS pictures as stimuli. As listed in Table 3.1, we defined the

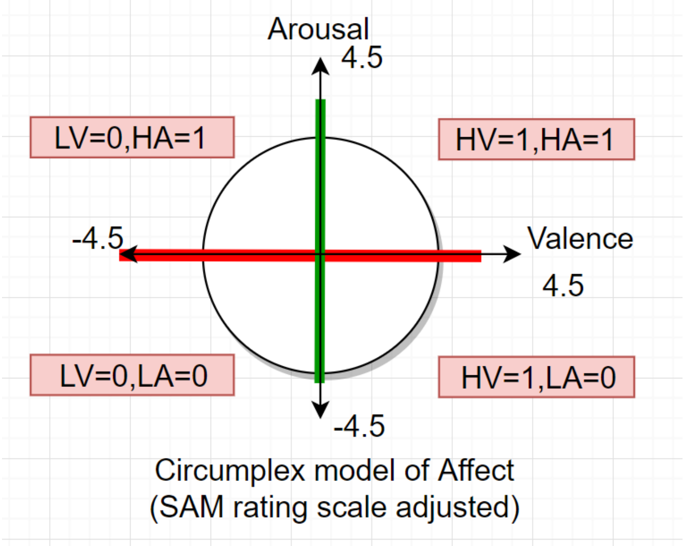


Figure 7.2: Four labels are used to mark the area of each quadrant in the Circumplex model of Affect. (HV=High Valence, LV=Low Valence, HA=High Arousal, LA=Low Arousal)

list of IAPS pictures to use to span 10 types of mood. Half of the samples are for the purpose of evoking displeasure (low valence) and the other half is for stimulating pleasure (high valence). The plot of the histogram of the labels of samples used to train the valence predictive model is displayed in 7.3. This figure shows that the two classes are balanced in the dataset chosen.

An observation that reinforces the appropriateness of the use of AUs for valence classification through predictive models should be mentioned here. One can refer to Table 2.1 (in Chapter 2) as a guide to map the relationship between AUs and emotion. Some AUs, such as AU9 (Nose wrinkler), can be highly indicative of disgust, directly. This is because the activation of AU9 is always linked to displeasure. Conversely, there are some AUs that can be mapped directly to pleasure(positive valence). Thus, AU vectors are well suited to be used as features to classify pleasure and displeasure.

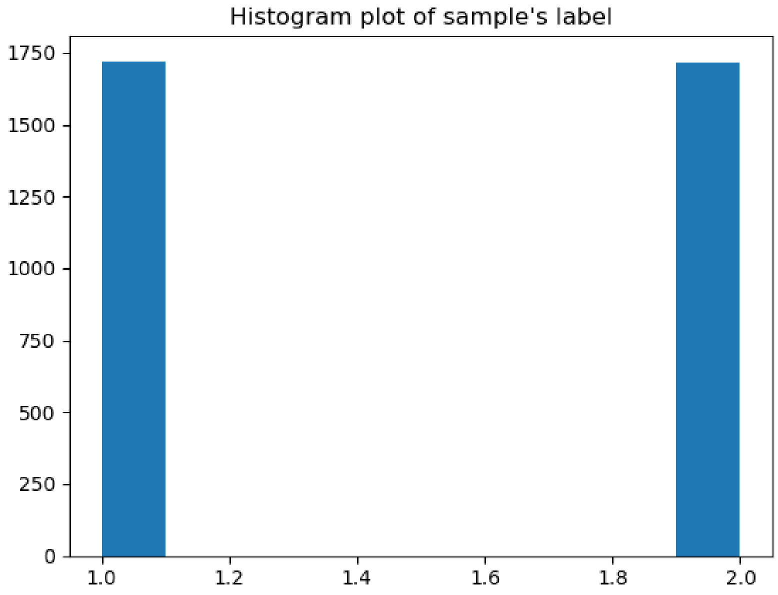


Figure 7.3: Histogram plot of samples' label (HV=1,LV=2)

7.4.2 Features for the Classification of Arousal

In the case of arousal, it is fortunate that there is a direct connection such that the pupillary response is proportional to the arousal level. However, the pupillary affective response is not the only factor that influences pupillary response. That is why we need to extract features that can distinguish the influence of the pupillary affect response without significant interference from the pupillary light reflex. In Chapter 6, we already discussed that we will use as classification features the attributes that we have discovered through our observation of the PD data. To prove that our features are suited to this task, we plotted the scatter plot matrix between the three features that we extract, which are ΔPQ, m_{QR}, and ΔQR, in Figure 7.4. Notice that points marked in light blue are labeled as LA (Low Arousal) and those in dark blue are labeled as HA (High Arousal). The two classes are aggregated in their own clusters, but one can see that they are linearly separable in the projections drawn. This type of problem is well suited for the use of an SVM as the classification algorithm.

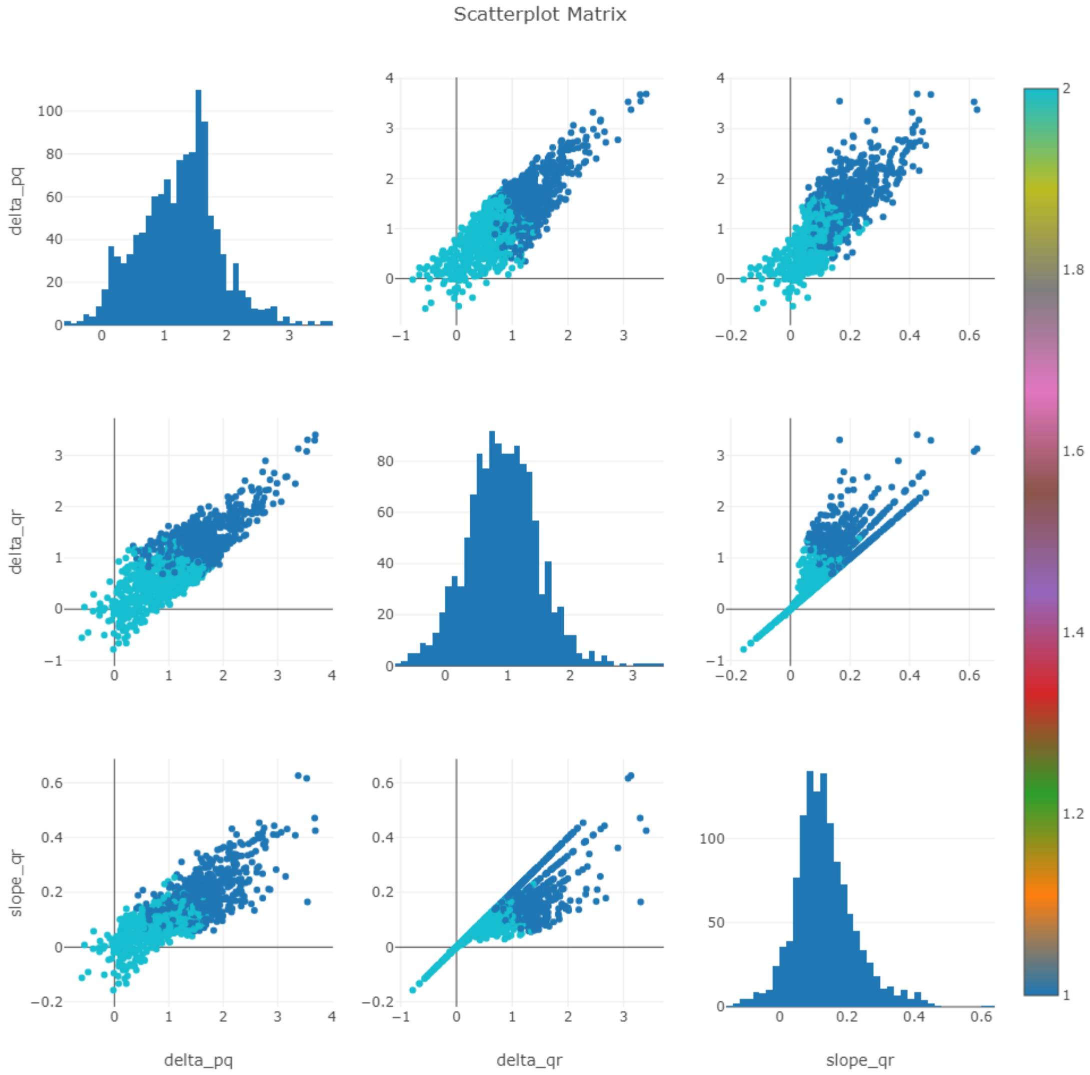


Figure 7.4: Scatter plot matrix between ΔPQ, m_{QR}, and ΔQR

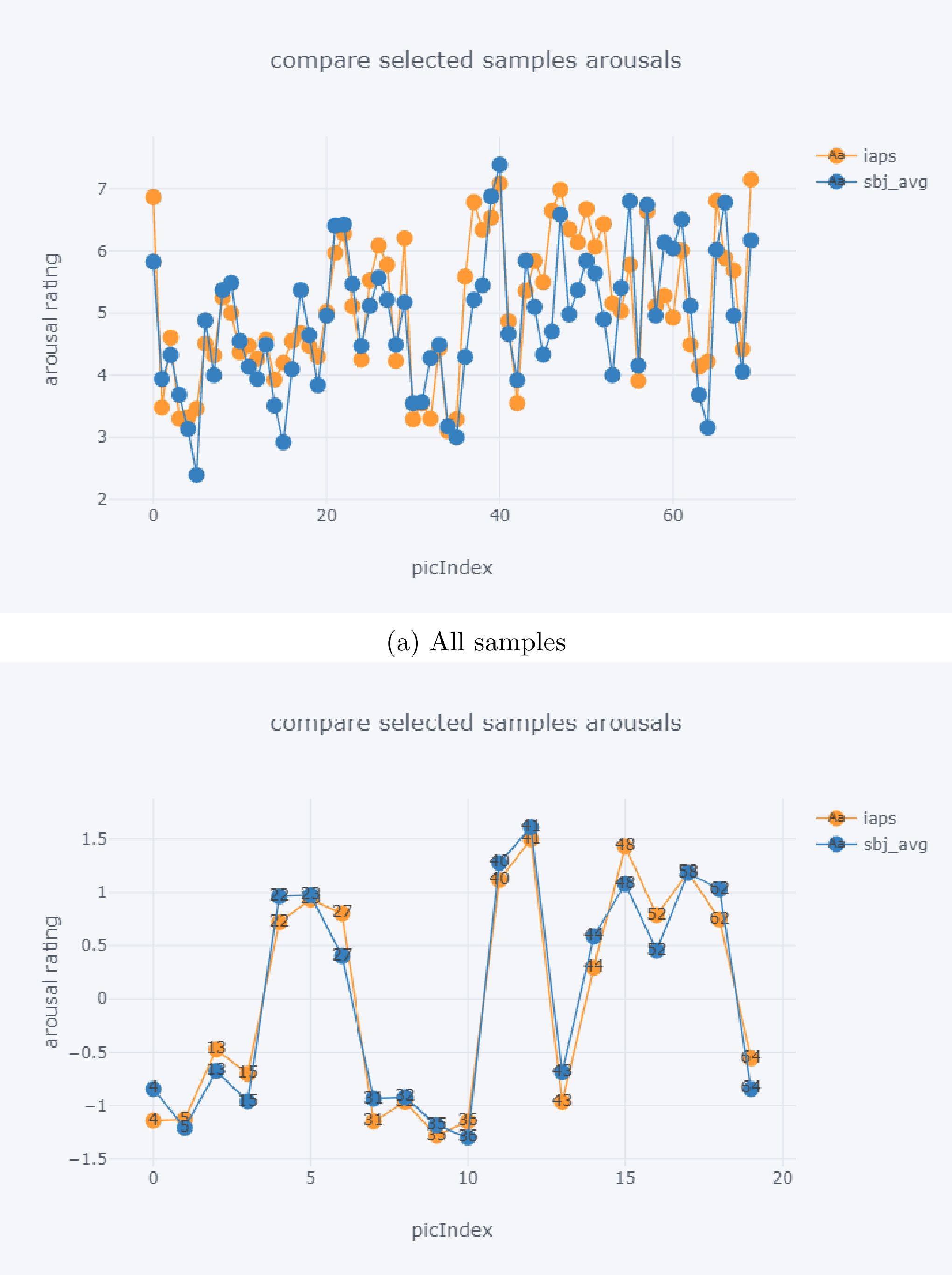
7.5 Data Selection

The selection of the data used for model training is one of the most important steps in building a predictive model. The model will learn according to the data provided to it during training, and this, therefore, has the potential to significantly impact the performance of the model.

In this respect, our data set presents ambiguity in the labels or target values (both for arousal and valence) that should be considered associated with each stimulus picture. We have, actually, two values of arousal and two values of valence that could be used as targets for each IAPS pictures. In each case, one is the mean value of the attribute provided along with the picture in the IAPS repository. The other one is the mean we have calculated from all the individual self-assessments marked (in the SAM tool) by our 50 subjects for that attribute of a given picture.

This kind of ambiguity is particularly worrisome for the arousal attribute, as it encodes the level of intensity in which the subjects experience emotion as a result of viewing an IAPS picture. Accordingly, we investigated this ambiguity further, comparing the two sets of arousal label values for all the 70 IAPS pictures displayed to the experimental subjects. The plot comparing between those two sets is shown in Figure 7.5a. In the figure, we observe that the two arousal labels for some pictures are far apart from each other, while for some pictures both arousal label values are similar.

To avoid this ambiguity issue, we have chosen to train the model with data obtained from the presentation of pictures that have both their arousal labels in close proximity. Figure 7.5b shows the plots comparing both arousal labels for the subset of pictures chosen. While this reduces the size of the training set, the decision was made with the aim to provide a consistent training set for the development of the predictive model, removing possible ambiguities that could act as confounding factors.



(a) All samples

(b) Selected samples

Figure 7.5: Plots of average SAM rating obtained in our study compare to the one from IAPS repository

7.6 Machine Learning Pipelines

The implementation programming code for the development of the predictive model was written in Python. We supported the development with the use of several popular 3rd party libraries for data science projects, such as Pandas, Numpy, Scipy, etc. In the implementation of the machine learning model and its pipeline, we used the Python library called Sklearn [PVG+11] which is well-known and widely used in both academia and industry. The specific model used for this study is developed with the Support Vector Classifier (SVC) toolbox which is part of the Sklearn library.

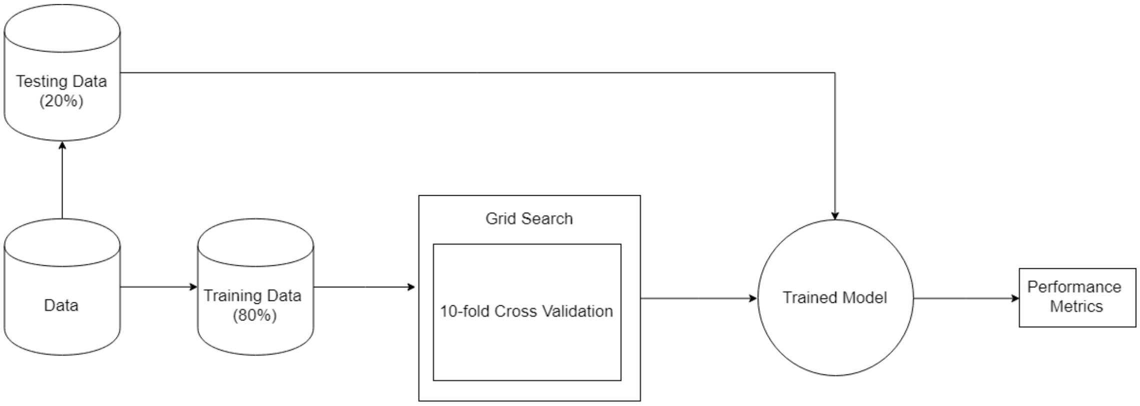


Figure 7.6: Machine learning pipelines for both valence and arousal classifiers

Our machine learning pipeline is summarized in the form of a diagram, shown in Figure 7.6. The data were subdivided into training and testing portions randomly. Then for each training session, we execute a 10-fold cross validation strategy to obtain a robust assessment of the performance of the model. Cross-validation is a common practice that has proved to help enhance the robustness of the model when dealing with a small dataset [MBD+90]. For tuning the hyperparameters of the learning process, we have conducted a grid search strategy and used accuracy and F1 score as metrics to choose the best model.

7.7 Summary

This chapter has explained the process followed for building the predictive model for this study, allowing the reader to develop a more intuitive understanding of how the model learns from the extracted features.

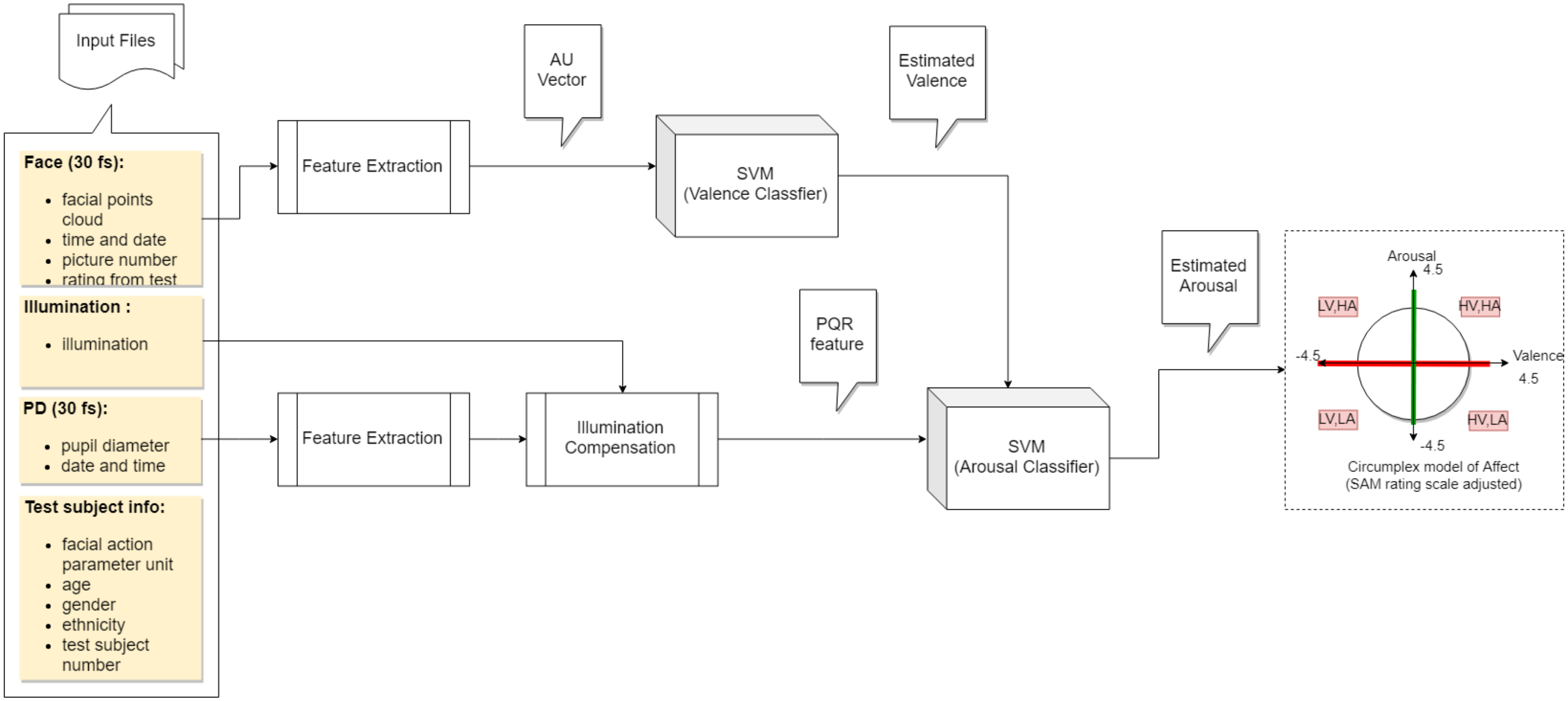


Figure 7.7: Diagram showing the process of how the predictive model is trained.

CHAPTER 8

RESULT AND DISCUSSION

This chapter presents the performance achieved by the predictive model. We will first show the performance of the valence classifier and arousal classifier when they are trained separately, and finally, the results obtained by the combined classifier. Additionally, we will compare the accuracy results of the models when illumination compensation is applied to cases in which illumination compensation was not applied.

Later in the chapter, we will discuss the significance and implications of the results along with the difficulties and limitations encountered along with the development of this system.

8.1 Result

Our hypothesis, which can be recalled from chapter 2, is that monitoring of the pupil diameter and facial expressions would enable the estimation of the valence and the arousal parameters in the Circumplex Model of Affect. In this study, we implemented a predictive model to determine the quadrant of the circumplex model in which the affective status of the subject is located. Each quadrant is marked based on their arousal and valence characteristics with four types of labels: HV (High Valence or 1), LV (Low Valence or 0), HA (High Arousal or 1), and LA (Low Arousal or 0). This was shown in Figure 7.2 (Chapter 7).

8.1.1 Valence Classifier

For estimating valence, we sought to classify between pleasure (HV) and displeasure (LV) based on the subjects facial expression. The proposed classifier SVM obtained

75% accuracy. Table 8.1 shows the confusion matrix of the valence classifier and Figure 8.1. displays its receiver operating characteristic (ROC).

Table 8.1: Confusion Matrix of Valence Classifier

	Predicted: HV	Predicted: LV
Actual: HV	487	127
Actual: LV	250	606

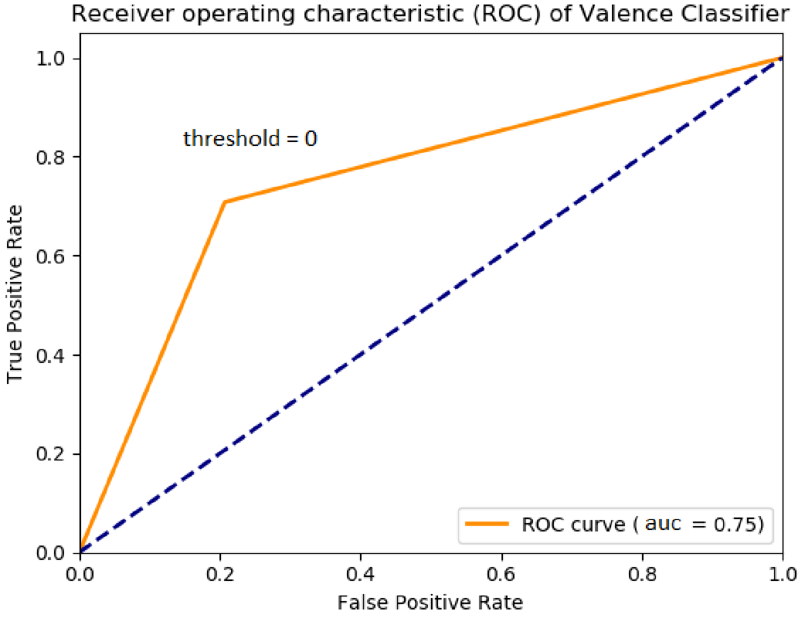


Figure 8.1: ROC plot of Arousal classifier

8.1.2 Arousal Classifier

In the case of arousal, the two classes that we try to classify is HA (High Arousal or 1) and LA (Low Arousal or 0). The results we obtained here are from training the arousal classifier without the valence feature. Table 8.2 shows the result of

84% accuracy without applying the illumination compensation method while the accuracy when illumination compensation was applied achieves 90% accuracy. The confusion matrices for both instances are shown in Tables 8.2 and 8.3. Their ROC plots can be observed in Figures 8.2a and 8.2b respectively.

Table 8.2: Confusion Matrix of Arousal classifier without illumination compensation

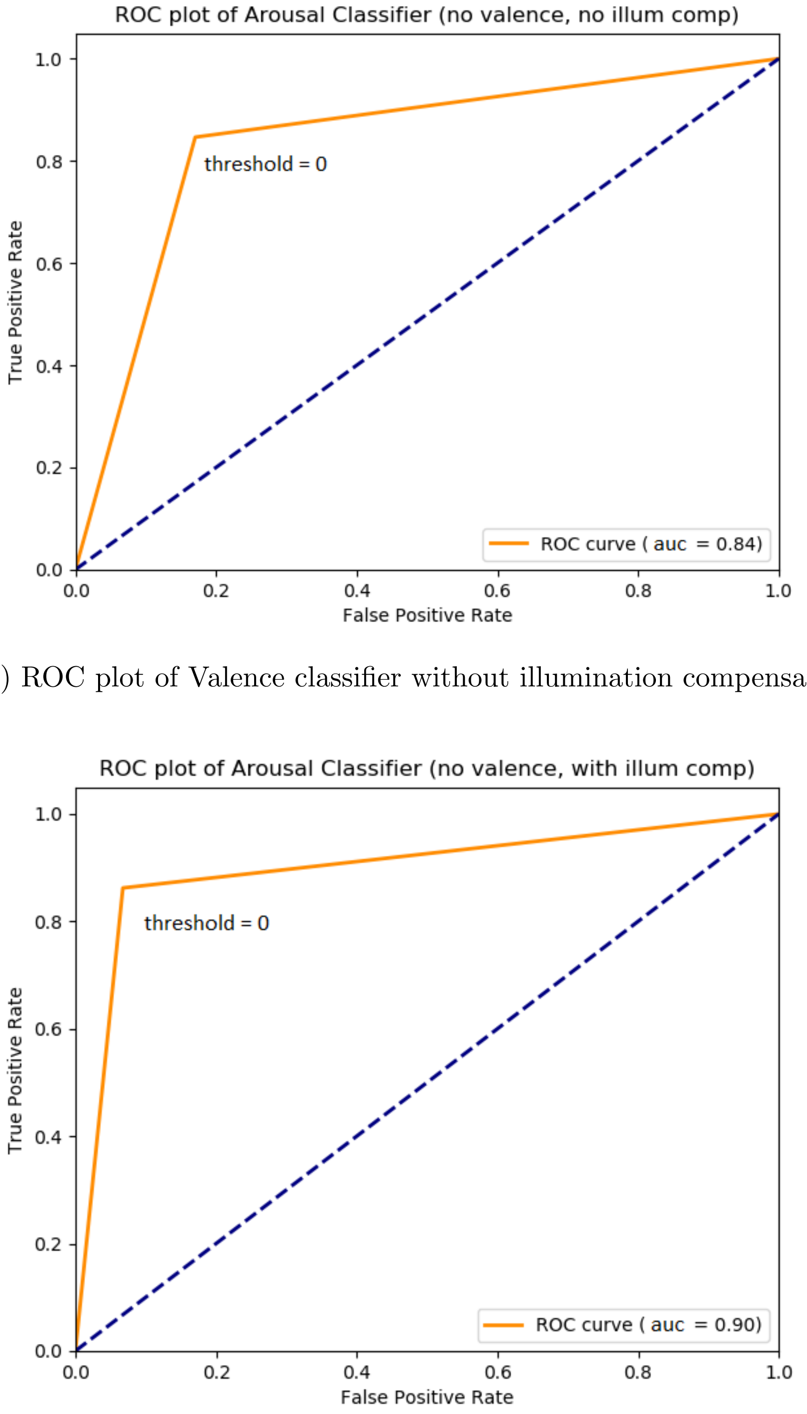
	Predicted: HA	Predicted: LA
Actual: HA	448	92
Actual: LA	85	468

Table 8.3: Confusion Matrix of Arousal classifier with illumination compensation

	Predicted: HA	Predicted: LA
Actual: HA	446	51
Actual: LA	43	457

8.1.3 Combined Classifier

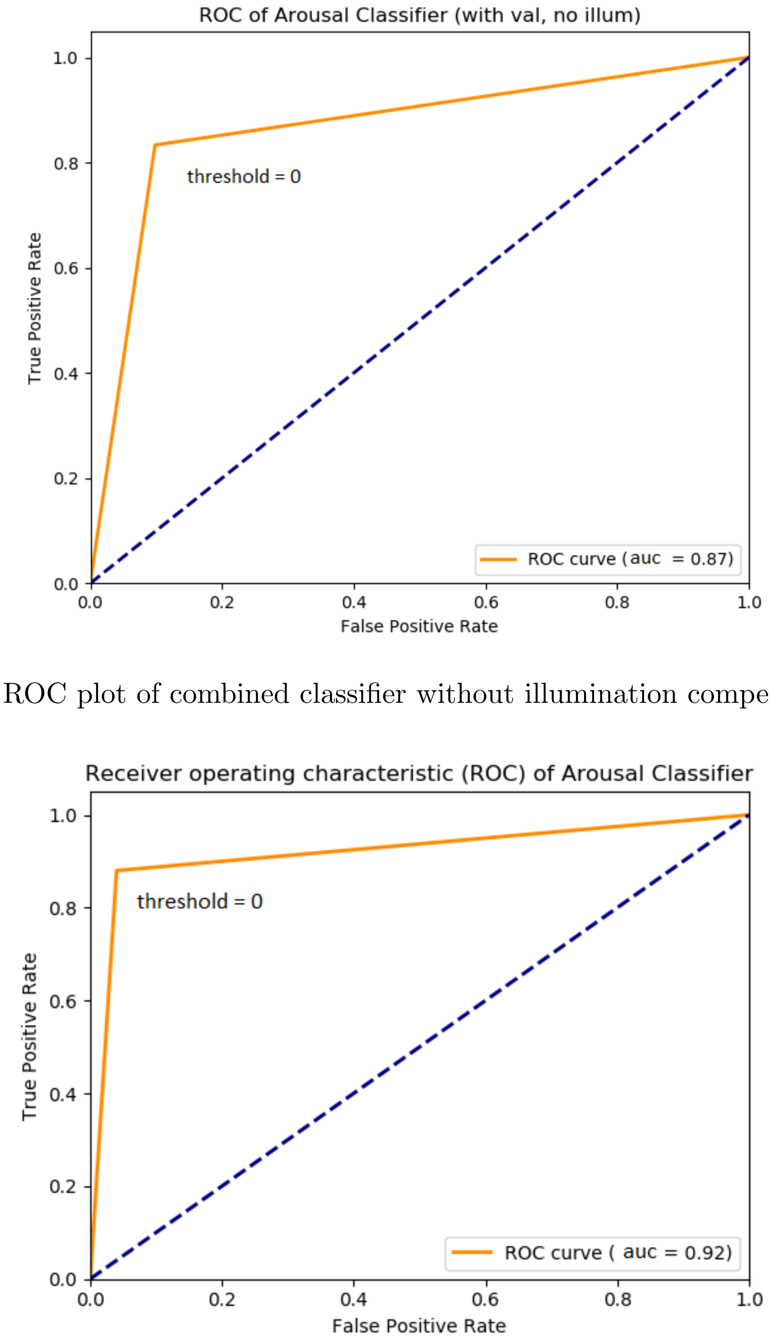
The combined classifier uses the result from the valence classifier, which is used as an input, along with the PQR features, to estimate the corresponding arousal level, therefore yielding the combined estimates of valence and arousal, which place the affective state of the subject in the Circumplex model. The accuracy of arousal estimations in the combined classifier with and without illumination compensation applied is 87% and 92%, respectively. The confusion matrices can be found in Tables 8.4 and 8.5, and the corresponding ROC plots appear in Figures 8.3a and 8.3b.



(a) ROC plot of Valence classifier without illumination compensation

(b) ROC plot of Arousal classifier with illumination compensation

Figure 8.2: Receiver operating characteristic (ROC)



(a) ROC plot of combined classifier without illumination compensation

(b) ROC plot of combined classifier with illumination compensation

Figure 8.3: Receiver operating characteristic (ROC)

Table 8.4: Confusion Matrix of Combined classifier without illumination compensation

	Predicted: HA	Predicted: LA
Actual: HA	55	6
Actual: LA	8	40

Table 8.5: Confusion Matrix of Combined classifier with illumination compensation

	Predicted: HA	Predicted: LA
Actual: HA	48	2
Actual: LA	6	44

8.2 Discussion

The results shown above confirm that it is viable to monitor the facial expression and pupil diameter of a subject to estimate the valence and arousal of a computer user, reaching an accuracy of 75% in valence and 92% in arousal. To better evaluate the performance of the approach proposed, it should be compared with other similar studies in this field of research. The 2015 paper A survey on Human Emotion Recognition approaches [VV15] provides a summary of similar developments in the field.

8.2.1 Results Comparison

In terms of valence classification, we can refer to Table 8.7, reproduced from [VV15]. Several approaches reported high accuracy in recognizing as many as 6 classes. Our approach resulted in a 75% accuracy for classifying two classes. However, several of the other methods reported are based on more complex sensing modalities and have heavier computational power requirements, which may complicate their deployment to everyday computer use applications. In that regard, our approach has the advantage of reducing the computational and memory requirements by encoding facial expressions into FAP vectors. Furthermore, none of those works used the self-reports through the SAM tool to define the labels for classification, relying on the characterizations provided in the stimuli databases (e.g., the IAPS database) for the definition of labels.

Taking that into account, the accuracy achieved in this study is a promising initial step in the direction of implementing affective valence sensing on the bases of very affordable, off-the-shelf instrumentation (e.g., the commercially available Kinect module), which may be more practical for broad adoption in the future.

In terms of arousal, we will refer to Table 8.8 provided by [VV15]. One study that is similar to ours is the study to classify the arousal level based on pupil diameter by P. Ren [RBGA12], which has 83.16% of accuracy. However, it should be noted that the study explored a narrower set of subject reactions, eliciting only two reactions in response to congruent and incongruent Stroop word presentations. In our study, we implement a novel approach to includes the (previously estimated) valence as one of the feature inputs to recognize the arousal level and achieves 92% accuracy as a result. This indicates an improvement in arousal recognition, achieved with the approach proposed in this dissertation

For convenience, we have summarized the accuracy results from the proposed system in Table 8.6.

Another factor that played an important role in the level achieved for arousal recognition in this study is illumination compensation. Its impact on the definition of the PQR features and on the performance of the model are discussed in the next section.

Table 8.6: Comparison of Model's accuracy in all cases

	Illumination Compensation	Accuracy
Valence	n/a	0.75
Arousal	No	0.84
Arousal	Yes	0.90
Combine	No	0.87
Combine	Yes	0.92

Table 8.7: Emotion recognition using facial expression. (Table from [VV15])

Work done by	Features used	Database & Recognition rate	No. of classes
Zisheg LI	I. Appearance II. Shape	C-K database, 96.33%	6
Ligang Zhang	Patch based 3D Gabor features	JAFFE,92.93% C-K database, 94.48%	6
Ramchand Hablani	Local Binary patterns	JAFFE, I. 94.44% II. 73.61%	6
Fadi Dornaika	3D Tracker based Facial actions	CMU database, Above 90%	6
Munawar Hayat	Local video patches	BU4DFE 94.34%	6
Seyedehsamaneh	Pose-invariant features based on Optical Flow	ECK+ Database I. 90.36 II. 92.74	6

Table 8.8: Emotion recognition using other modalities. (Table from [VV15])

Work Done By	Features used	Database & Recognition rate	No. of classes
P.Ren	Pupil Diameter Signal	Own Database, 83.16%	2
G.Caridakis	Facial features, Body gestures and Speech parameters	Own database, >70%	8
K.Tang	2D,3D features, speech features	Data from Kinect Sensor, >70%	7
M.Soleymani	EEG features, Pupilliary response & gaze distance	Own database, 68.5% for Valence 76.4% for Arousal	6
W.Zheng	EEG features, Pupilliary response	Own database, 73.59% 72.98%	3
J.Wagner	Vocal Features, Facial Expressions and Gestures	CALLAS Expressivity Corpus,–	4

8.2.2 Influence of Illumination Compensation

In addition to the modification of the pupil diameter in response to affective stimuli, illumination is known to significantly influence the aperture of the pupil. We have discussed how we compensate the illumination effect in Chapter 6 and now we will highlight the impact of that compensation on the performance of the classifier. Table 8.6 demonstrates how affect estimation results are significantly improved after applying the illumination compensation algorithm. We also plot the scatter matrix of PQR features before and after the compensation technique (See Figures 8.4 and 8.5). Comparing the two plots, we can see that the data clusters formed by the two classes are more separable after the illumination compensation technique is applied. This confirmed that our technique for compensating the illumination factor is fulfilling its purpose.

8.2.3 Challenges and Limitation

As the field of affective recognition continues to evolve, the process of proposing new approaches needs to be based on a number of assumptions which may only be verified partially during the experimental work of the research efforts. It is only after the completion of the experimental work and the analysis of the data collected that the assumptions can be revisited and reconsidered.

The discovery of unexpected factors that have been encountered along the way is part of the learning process. We have next listed all such factors that we have discovered in this section, in the hope that the future studies can use this as a guideline for areas where assumptions need to be carefully considered.

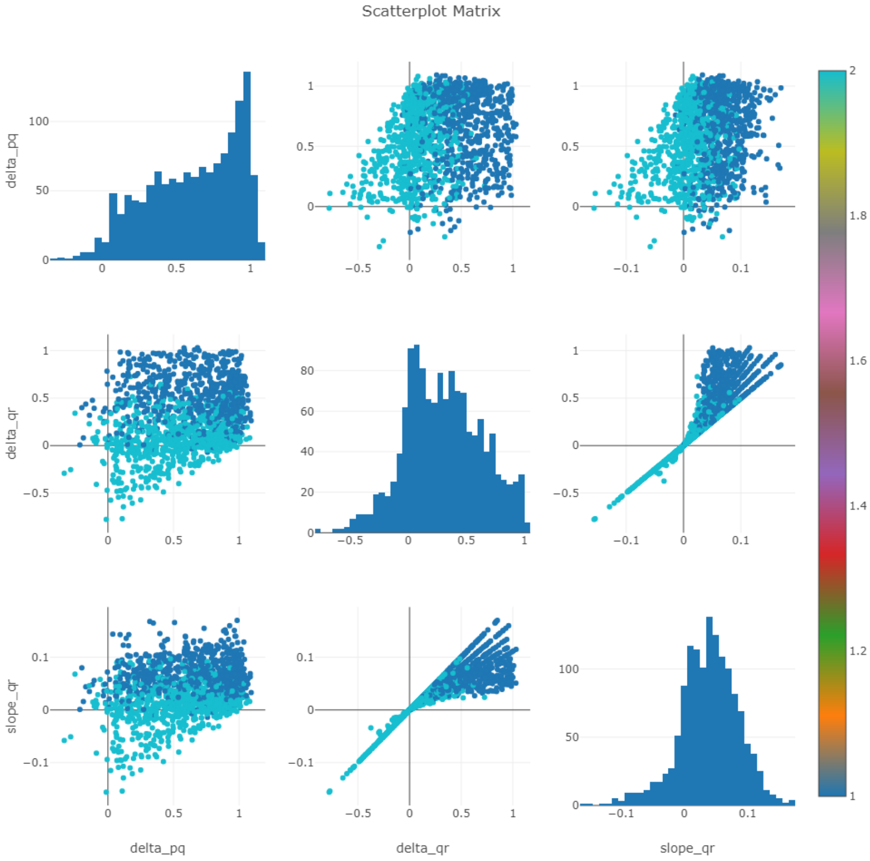


Figure 8.4: Scatter plot of PQR features without applying illumination compensation technique

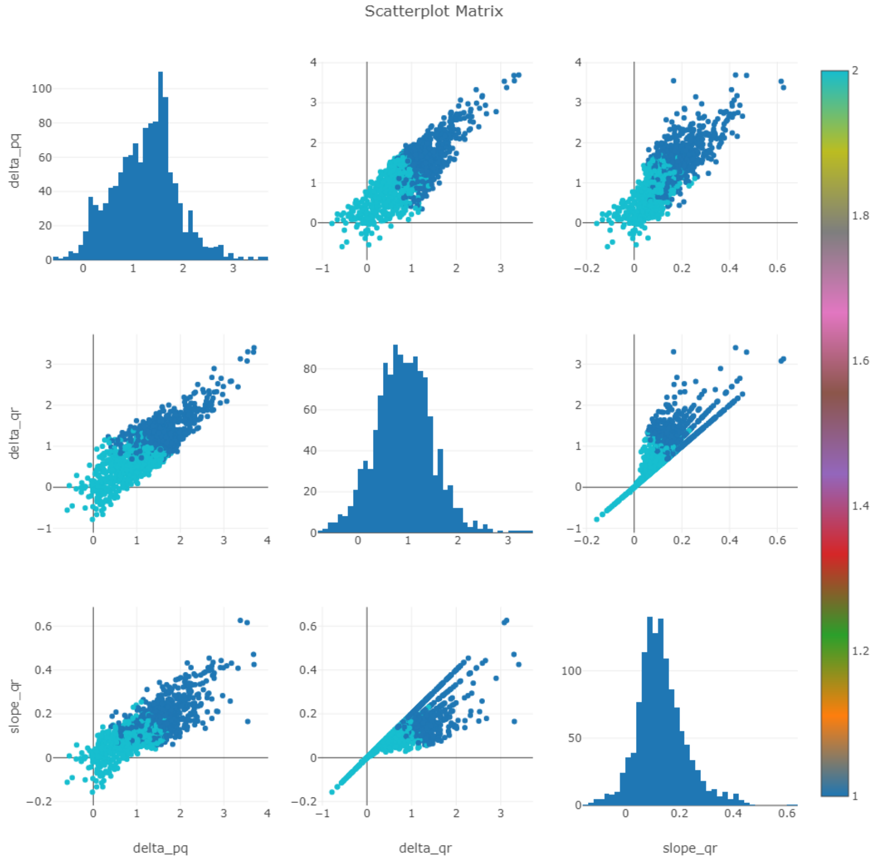


Figure 8.5: Scatter plot of PQR features applied by illumination compensation technique

Human Factor

The emotional mechanisms experienced by human beings are complex and frequently modulated by highly individual factors. This makes it difficult to attach a definitive label to the levels arousal or valence that can be expected as the response of a given experimental subject to the presentation of a pictorial stimulus (e,g., a picture from the IAPS database). As we have described, there are discrepancies between the plot of the expected arousal responses obtained as the means of arousal in the IAPS repository and the plot drawn from the averaged arousal rankings of the 50participants in our experiment (see Figure 7.5a). The plots show that these two sets of expected arousal responses are frequently not consistent with each other. There are many factors behind this observed effect. One factor that may be contributing to the discrepancy observed may relate to the different gender ratios in the subject populations that generated those arousal mean values. Participants of different genders may exhibit heightened arousal in response to different pictorial themes. Also, we observed that the self-awareness of each participant regarding the level of arousal reached may vary from individual to individual, which could cause further ambiguity in the definition of arousal labels for the pictures. Another factor possibly contributing the difficulty in establishing robust arousal labels might be the impact of the cultural background of the individuals on their reaction to the images used. These were some of the main reasons why the decision was made to only use IAPS images that had consistency in their expected arousal levels as defined from both information sources.

Depth Intolerance

Although our experimental subjects were asked to remain as still as possible, position shifts and adjustments were frequently observed through the experiment (which

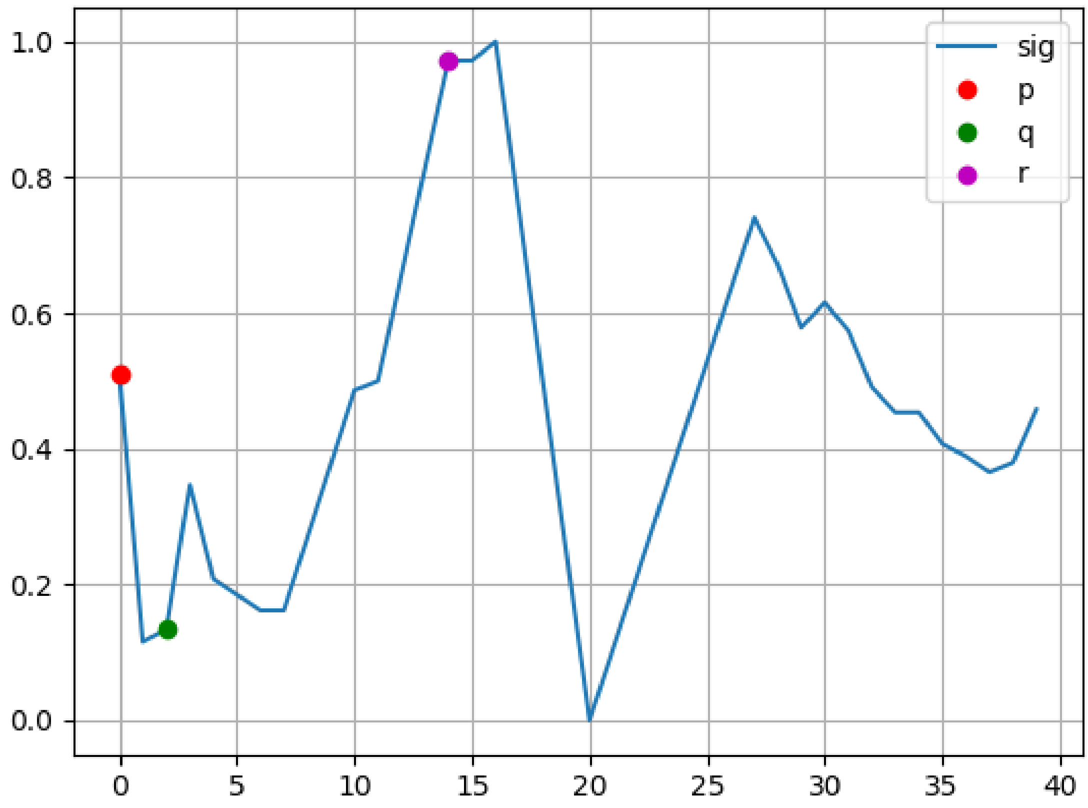


Figure 8.6: Plot of one sample of pupil diameter signal that is corrupted by the test subject's movement

lasted about 45 minutes). These movements introduced artifacts and scaling factors in both the pupil diameter and the face surface that caused some parts of the recorded signals to be discarded. For example, one sample of pupil diameter signal is shown in Figure 8.6. The artifacts introduced in this pupil diameter record, for example, resulted in having to discard it, because the relative sizes of ΔQR and ΔPQ in the sample are abnormal, likely due to the introduction of depth variations.

Device Limitation

The eye-gaze tracker device used in this study is attractive because of its portability and size. However, it does not offer very high resolution in the pupil diameter measurements that are obtained from it. This may have been part of the reason for the presence of noise in the pupil diameter signals recorded. As described in Chapter 6, digital signal processing techniques were utilized to try to minimize the impact of that noise in the feature extraction process, but those are no substitute for a cleaner original signal, which would be highly desirable.

Ethical Limitation

We relied on the expectation that the stimulus pictures from the IAPS database would effectively elicit perceivable emotional responses in our experimental subjects. However, in deciding which subset of the IAPS images to use for the experiment the desire to potentially use images with the highest arousal mean values within the database, to provide powerful stimuli and prospectively elicit strong responses, had to be counterbalanced by the commitment to the safety and well being of the experimental subjects. Accordingly, this study did not use some images that had high arousal levels in the IAPS database but could have resulted too shocking for our experimental subjects (for example, headless body). Preserving this level of moderation in the strength of the stimuli may have resulted in an inability to fully explore the spectrum of emotional reactions of the subjects, but that was a conscious decision made for the reasons stated above.

CHAPTER 9

CONCLUSION AND FUTURE WORK

This final chapter will provide a retrospective summary of the development of this project, emphasizing the conclusions that can be reached on the basis of the experimental results and their analysis. This chapter will also reflect on suggestions for future development and improvement that emerge from the outcomes of this dissertation.

9.1 Conclusion

The motivation for the research reported in this dissertation was the definition of an unobtrusive approach for the assessment of the affective status of a computer user. The Circumplex Model of Affect, with its dimensions of arousal and valence, was chosen as the frame of reference for the estimation of the users affective state. Further, it was decided that the affective assessment of the user would be attempted on the basis of information extracted from the pupil diameter, monitored with an eye-gaze tracker and the facial expression, monitored using a Kinect module. The information obtained from the user would be processed following a machine learning approach to yield estimates of the affective arousal and affective valence levels experienced by the computer user.

The pupil diameter is known to be an indicator of arousal in humans since the pupil size is influenced by Autonomic Nervous System (ANS) which controls the arousal level. Unfortunately, the pupillary response is also affected by the pupillary light reflex which is the bodys mechanism to regulate the amount of light that reaches the retina. To address this issue, specific modifications, controlled by the estimated illumination level, were inserted in the feature extraction process applied to the measured pupil diameter signal.

As it is speculated that the pupil diameter is not reflective of the valence of the affective state of the subject, it was necessary to extract this parameter from the users facial expression, monitored using the Kinect module. In particular, this dissertation proposed the embedding of the facial expression information into Facial Animation Parameter (FAP) vectors to detect the activation of Action Units (AU) that can be used as features for affective valence estimation. The FAP vectors have the added attribute to be face-size independent, as they are unitless.

In the pursuit of the goal mentioned above, the AffectiveMonitor system was developed, which involves to hardware sub-systems (Primary Side and Secondary Side) and the software that controls both of the sub-systems. The AffectiveMonitor system controls both instruments used for data acquisition (eye gaze tracker and Kinect module) and performs all the pre-processing and feature extraction tasks, following the procedures detailed in this dissertation. Furthermore, this system has multiple modes of operation. One of its modes allows the recording of features derived from pupil diameter and facial expression measurements gathered while experimental subjects are presented with pictures from the International Affective Picture Systems (IAPS) database that are meant to shift the viewers affective state to arousal and valence levels assessed in previous experiments performed by the IAPS developers. This mode of the AffectiveMonitor system was used to collect data for the development of the machine learning predictive model for affective assessment.

The contribution of this research work extends beyond the development of the AffectiveMonitor platform into the adaptation of techniques and algorithms to accommodate the processing challenges encountered in the pursuit of the affective assessment goal.

It was found, for example, that in multiple instances there are significant discrepancies between the mean values of arousal and valence published with each IAPS image and the averages obtained through self-reporting by the 50 experimental subjects enrolled in the data collection process completed for this dissertation (with FIU Institutional Review Board approval). It was then necessary to set discard data collected from presentation of images with large discrepancies in their arousal mean levels.

Similarly, an alternative procedure for illumination compensation (embedded in the feature extraction stage) had to be devised when it became apparent that the illumination compensation process used in previous research from the FIU DSP Laboratory was not practical for application to data collected in short intervals (10 seconds) as recommended by the authors of the IAPS database.

This research also addressed the challenge of locating the changes in pupil diameter and in facial expression to specific intervals within the 10-second recording window that followed the presentation of each one of the IAPS images to the subjects. The automated procedure devised for this purpose was necessary to account for the variable latency at which facial expression changes occur for different subjects.

The machine learning model developed as a result of this dissertation used the Support Vector Machine (SVM) architecture in a cascade configuration which estimates the affective valence first, based on Kinect data and supplements the assessment with an estimation of the arousal level in a second step. 10-fold cross validation processes yielded estimation accuracies of 75% for the assessment of valence level and 92% for the assessment of arousal level.

It was also observed that the inclusion of the illumination compensation in the feature extraction process for the pupil diameter signals played, as expected, an important role in enhancing the arousal accuracy recognition by compensating the effect from the Pupillary Light Reflex (PLR).

Overall, this work has completed the development of the hardware and software integration of a novel non-intrusive system for the automated assessment of a computer users affective state (in terms of valence and arousal), using an off-the-shelf portable eye tracking module and a standard Microsoft Kinect module. The development and verification of this original system created on the basis of ordinary off-the-shelf sensors has provided a new avenue for the prospective development of affective sensing systems that could be practically deployed to ordinary computer setups, as they will not involve intrusive interactions with the users or the requirement for highly specialized and elaborate sensing instruments that are not usually available to the ordinary computer user.

9.2 Future Work

In its current state, the system developed in this study, classifies the affective arousal and valence of the user to place it in one of the quadrants of the Circumplex Model of Affect. Future developments will likely pursue the classification of each one of the attributes (arousal and valence) with a finer granularity, to be able to locate the affective state of the computer user to more specific regions within the circumplex model (see Figure 9.1).

As the potential benefits of Deep Learning approaches in machine learning solutions become better understood, it is tempting to envision the development of an affective assessment system like the one developed in this dissertation along the lines of Deep Learning principles. In particular, this would imply that the first layers of the Deep Learning system could, to some extent, define, in a data-driven fashion

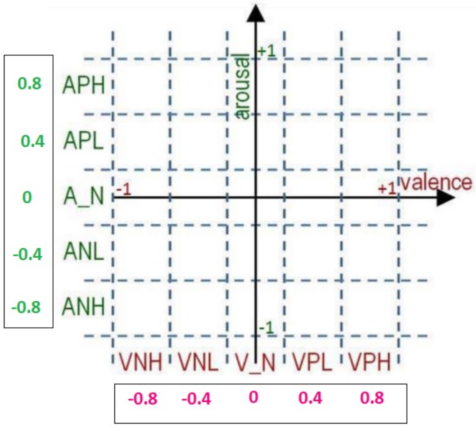


Figure 9.1: Fine-scale regions in the Circumplex model of affect

the best feature extraction approach, operating directly on aggregates of raw data or only lightly-reduced data. This prospect, however, has to be considered in the context of the much larger amounts of training data that are usually necessary to build Deep Learning models. In the arena of affective assessment this amounts to a daunting data collection process that would probably have to involve large numbers of human subjects completing very specific experimental protocols.

Lastly, it may be of interest to explore the monitoring of the pupil diameter signal through the use of compact and affordable eye gaze tracking models that are now emerging in the market as a consequence of the significant advances in high-definition digital camera modules. Some of those devices could be considered to propose an even less specialized and costly set of instruments for an affective assessment system. However, the affordability and compactness of the modules chosen should always be secondary to the need to have high resolution and low noise levels in pupil diameter signals collected.

BIBLIOGRAPHY

[AA01] Jörgen Ahlberg and Jörgen Ahlberg. CANDIDE-3 - An Updated Parameterised Face. 2001.

[AMu] J Seymour Graphic Arts, Proceedings of the 61st Annual Meeting, and undefined 2009. Color measurement with an RGB camera. *researchgate.net*.

[AOC03] Valitutti Alessandro, Stock Oliviero, and Strapparava Carlo. Developing affective lexical resources. 2003.

[Bar08] Armando Barreto. Non-intrusive physiological monitoring for affective sensing of computer users. In *Human Computer Interaction: New Developments*. InTech, 2008.

[BL94] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, mar 1994.

[BMEL08] Margaret M. Bradley, Laura Miccoli, Miguel A. Escrig, and Peter J. Lang. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607, 2008.

[Bre12] Eli Bressert. *SciPy and NumPy: an overview for developers.* " O'Reilly Media, Inc.", 2012.

[BrEu] PC Bressloff, CV Wood Physical review E, and undefined 1998. Spontaneous oscillations in a nonlinear delayed-feedback shunting model of the pupil light reflex. *APS*.

[BW98] PC Bressloff and CV Wood. Spontaneous oscillations in a nonlinear delayed-feedback shunting model of the pupil light reflex. *Physical review E*, 58(3):3597, 1998.

[BZRG07] Armando Barreto, Jing Zhai, Naphtali Rishe, and Ying Gao. Measurement of pupil diameter variations as a physiological indicator of the affective state in a computer user. *Biomedical sciences instrumentation*, 43:146–151, 2007.

[CDCT$^+$01] Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80, 2001.

[CHFT06] Ya Chang, Changbo Hu, Rogerio Feris, and Matthew Turk. Manifold based analysis of facial expression. *Image and Vision Computing*, 24(6):605–614, 2006.

[COG11] COGAIN: Communication by Gaze Interaction. Eye Tracker Eyetech - COGAIN: Communication by Gaze Interaction (hosted by the COGAIN Association), 2011.

[Dam05] Antonio Damasio. The neurobiological grounding of human values. In *Neurobiology of human values*, pages 47–56. Springer, 2005.

[Duc] Andrew T Duchowski Theory. Eye Tracking Methodology. Technical report.

[EFA80] Paul Ekman, Wallace V Freisen, and Sonia Ancoli. Facial signs of emotional experience. *Journal of personality and social psychology*, 39(6):1125, 1980.

[Ell] C Elliot. The affective reasoner: A process model of emotions in a multi-agent system. 1992. *Northwestern University Institute for the Learning Sciences: Northwestern, IL*, 48.

[GBA09a] Ying Gao, Armando Barreto, and Malek Adjouadi. Detection of Sympathetic Activation through Measurement and Adaptive Processing of the Pupil Diameter for Affective Assessment of Computer Users. *American Journal of Biomedical Sciences*, 1(4):283–294, 2009.

[GBA09b] Ying Gao, Armando Barreto, and Malek Adjouadi. Detection of sympathetic activation through measurement and adaptive processing of the pupil diameter for affective assessment of computer users. *Am. J. Biomed. Sci*, 1(4):283–294, 2009.

[GBSu] Y Gao, A Barreto, M Adjouadi Am. J. Biomed. Sci, and undefined 2009. Detection of Sympathetic Activation through Measurement and Adaptive Processing of the Pupil Diameter for Affective Assessment of Computer Users. *pdfs.semanticscholar.org*.

[GD05] Guodong Guo and Charles R Dyer. Learning from examples in the small sample case: face expression recognition. *IEEE Transactions on*

Systems, Man, and Cybernetics, Part B (Cybernetics), 35(3):477–488, 2005.

[GSH$^+$00] Ben Goertzel, Ken Silverman, Cate Hartley, Stephan Bugaj, and Mike Ross. The baby webmind project. In *Proceedings of AISB*, 2000.

[HE11] Peter D Hiscocks and P Eng. Measuring luminance with a digital camera. *Syscomp electronic design limited*, 16, 2011.

[Hud03] Eva Hudlicka. To feel or not to feel: The role of affect in human–computer interaction. *International journal of human-computer studies*, 59(1-2):1–32, 2003.

[Hug95] Kenneth Hugdahl. *Psychophysiology: The mind-body perspective.* Harvard University Press, 1995.

[Jus09] Justin Weaver. quicklinkapi4net, 2009.

[KSS18] Mariska E Kret and Elio E Sjak-Shie. Preprocessing pupil size data: Guidelines and code. *Behavior research methods*, pages 1–7, 2018.

[Kui99] JB Kuipers. Quaternions and rotation sequences: a primer with applications to orbits, aerospace and virtual reality, kuipers jb, princeton university press, 41 william street, princeton, nj 08540, usa. 1999. 372pp. *The Aeronautical Journal*, 1999.

[LBC99] PJ Lang, MM Bradley, and BN Culthbert. International affective digitized sounds (iads): Stimuli, instruction manual and affective ratings (tech. rep. no. b-2). *The Center for Research in Psychophysiology, University of Florida, USA*, 83, 1999.

[LLS03] Hugo Liu, Henry Lieberman, and Ted Selker. A model of textual affect sensing using real-world knowledge. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 125–132. ACM, 2003.

[Low] Phillip Low. Overview of the Autonomic Nervous System - Brain, Spinal Cord, and Nerve Disorders - Merck Manuals Consumer Version.

[LP05] LANG and PJ. International affective picture system (IAPS) : affective ratings of pictures and instruction manual. *Technical Report*, 2005.

[Mat01a] David Matsumoto. Culture and emotion. *The handbook of culture and psychology*, pages 171–194, 2001.

[Mat01b] David Matsumoto. *The handbook of culture and psychology.* Oxford University Press, 2001.

[MBD$^+$90] Tom Mitchell, Bruce Buchanan, Gerald DeJong, Thomas Dietterich, Paul Rosenbloom, and Alex Waibel. Machine learning. *Annual review of computer science*, 4(1):417–433, 1990.

[nas] Apollo IMU Gimbal Lock.

[Neg18] Lucas Hermann Negri. Peakutils. *URL: https://pypi. python. org/pypi/PeakUtils (visited on 07/20/2016)*, 2018.

[NN56] Vincent Nowlis and Helen H Nowlis. The description and analysis of mood. *Annals of the New York Academy of Sciences*, 65(4):345–355, 1956.

[P$^+$95] Rosalind Wright Picard et al. Affective computing. 1995.

[PF03] Igor S Pandzic and Robert Forchheimer. *MPEG-4 facial animation: the standard, implementation and applications.* John Wiley & Sons, 2003.

[Pic99] Rosalind W Picard. Affective computing for hci. In *HCI (1)*, pages 829–833. Citeseer, 1999.

[Pic03] Rosalind W Picard. Affective computing: challenges. *International Journal of Human-Computer Studies*, 59(1-2):55–64, 2003.

[PS03] Timo Partala and Veikko Surakka. Pupil size variation as an indication of affective processing. *International journal of human-computer studies*, 59(1-2):185–198, 2003.

[PVG$^+$11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[Rah17] Mansib Rahman. Understanding How the Kinect Works. In *Beginning Microsoft Kinect for Windows SDK 2.0*, pages 21–40. Apress, Berkeley, CA, 2017.

[RBGA12] Peng Ren, Armando Barreto, Ying Gao, and Malek Adjouadi. Affective assessment by digital processing of the pupil diameter. *IEEE Transactions on Affective computing*, 4(1):2–14, 2012.

[Riz15] Donald C Rizzo. *Fundamentals of anatomy and physiology.* Cengage Learning, 2015.

[Rus80] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[S$^+$11] Ronald W Schafer et al. What is a savitzky-golay filter. *IEEE Signal processing magazine*, 28(4):111–117, 2011.

[SK88] Robert D. Strum and Donald E. Kirk. *First principles of discrete systems and digital signal processing.* Addison-Wesley, 1988.

[SSCP04] Stuart R Steinhauer, Greg J Siegle, Ruth Condray, and Misha Pless. Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International journal of psychophysiology*, 52(1):77–86, 2004.

[TJ18] Lizhe Tan and Jean Jiang. *Digital signal processing: fundamentals and applications.* Academic Press, 2018.

[Tob15] TobiiPro.com Learning Center. Dark and bright pupil tracking, 2015.

[TOlR$^+$18] Sudarat Tangnimitchok, Nonnarit O-larnnithipong, Neeranut Ratchatanantakit, Armando Barreto, Francisco R. Ortega, and Naphtali D. Rishe. A System for Non-intrusive Affective Assessment in the Circumplex Model from Pupil Diameter and Facial Expression Monitoring. pages 465–477. Springer, Cham, jul 2018.

[VV15] C Vinola and K Vimaladevi. A survey on human emotion recognition approaches, databases and applications. *ELCVIA: electronic letters on computer vision and image analysis*, pages 00024–44, 2015.

[YRS11] Michelle Yik, James A. Russell, and James H. Steiger. A 12-Point Circumplex Structure of Core Affect. *Emotion*, 11(4):705–731, 2011.

[ZJZY08] Yongmian Zhang, Qiang Ji, Zhiwei Zhu, and Beifang Yi. Dynamic facial expression analysis and synthesis with mpeg-4 facial animation parameters. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(10):1383–1396, 2008.

[ZPRH09] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, 31(1):39–58, 2009.

VITA

SUDARAT TANGNIMITCHOK

January 22, 1989	Born, Bangkok, Thailand
2012	B.E., Mechatronic Engineering Assumption University Bangkok, Thailand
2012	Research Intern Haute Ecole d'Ingénierie et de Gestion du Canton de Vaud (HEIG-VD) Yver-don-les-bain, Switzerland
2013–2014	Hardware and Software Engineering Powerline LLC (Tiptop Audio) Bangkok, Thailand
2017	Ph.D. Candidate Florida International University Miami, Florida

PUBLICATIONS AND PRESENTATIONS

Tangnimitchok S., Barreto A., O-larnnithipong N., Ratchatanantakit N., Ortega F.R., Rishe N.D.. *A System for Non-Intrusive Affective Assessment in the Circumplex Model from Pupil Diameter and Facial Expression Monitoring* . Manuscript submitted for publication in: Human-Computer Interaction. Interaction Platforms and Techniques. HCI 2018.

O-larnnithipong N., Tangnimitchok S., Barreto A., Ratchatanantakit N., Ortega F.R., *Real-Time Implementation of Orientation Correction Algorithm for 3D Hand Motion Tracking Interface* . Manuscript submitted for publication in: Human-Computer Interaction. Interaction Platforms and Techniques. HCI 2018. (accepted)

O-larnnithipong N., Tangnimitchok S., Barreto A., Ratchatanantakit N., *Orienta-*

tion Correction for a 3D Hand Motion Tracking Interface using Inertial Measurement Units . Manuscript submitted for publication in: Human-Computer Interaction. Interaction Platforms and Techniques. HCI 2018. (accepted)

Tangnimitchok S., O-larnnithipong N., Barreto A., Ortega F.R., Rishe N.D. *Finding an Efficient Threshold for Fixation Detection in Eye Gaze Tracking* . In: Kurosu M. (eds) Human-Computer Interaction. Interaction Platforms and Techniques. HCI 2016. Lecture Notes in Computer Science, vol 9732. Springer, Cham., pp. 93-103

O-larnnithipong N., Tangnimitchok S. and Barreto Armando, *Gyroscope Drift Correction Algorithm for Inertial Measurement Unit Applications in Robotics* , Proc. 2016 Florida Conference on Recent Advances in Robotics (FCRAR), May 12 13, 2016, Miami, FL., pp. 32 37.

Abyarjoo, F., Nonnarit, O., Tangnimitchok, S., Ortega, F. and Barreto, A., 2015. *PostureMonitor: Real-Time IMU Wearable Technology to Foster Poise and Health* . In Design, User Experience, and Usability: Interactive Experience Design (pp. 543-552). Springer International Publishing.

Janwattanapong, P., Ratchatanantakit, N., Tangnimitchok, S., O-larnnithipong, N., Aphiratsakun, N., *Implementation and Design of the AU Self-Balancing Bicycle (AUSB)* , The ECTI Conference on Application Research and Development (CARD) 2012, 21-22 June 2012, Pathumthani, Thailand.

Francisco R. Ortega, Fatemeh Abyarjoo, Armando Barreto, Napthali Rishe, Malek Adjouadi. *Interaction Design for 3D User Interfaces: The World of Modern Input Devices for Research, Applications, and Game Development* . Creating Home-Brew Devices with Arduino Microcontrollers, Chapter 24. CRC Press. December 2015.