AN APPROACH FOR THE EXTRACTION OF THERMAL FACIAL SIGNATURES FOR EVALUATING THREAT AND CHALLENGE EMOTIONAL STATES

Dissertation

Submitted to

The School of Engineering of the UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for

The Degree of

Doctor of Philosophy in Electrical Engineering

By

Nilesh Powar

Dayton, Ohio

December, 2013
AN APPROACH FOR THE EXTRACTION OF THERMAL FACIAL SIGNATURES FOR EVALUATING THREAT AND CHALLENGE EMOTIONAL STATES

Name: Powar, Nilesh

APPROVED BY:

Vijayan K. Asari, Ph.D.  
Advisory Committee Chairman  
Professor  
Dept. of Electrical and Computer Engineering

Russell C. Hardie, Ph. D.  
Committee member  
Professor  
Dept. of Electrical and Computer Engineering

John S. Loomis, Ph.D  
Committee Member  
Associate Professor  
Dept. of Electrical and Computer Engineering

Erik Blasch, Ph.D.  
Committee Member  
Air Force Research Laboratory

John G. Weber, Ph.D.  
Associate Dean  
School of Engineering

Tony E. Saliba, Ph.D.  
Dean, School of Engineering  
& Wilke Distinguished Professor
ABSTRACT

AN APPROACH FOR THE EXTRACTION OF THERMAL FACIAL SIGNATURES FOR EVALUATING THREAT AND CHALLENGE EMOTIONAL STATES

Name: Powar, Nilesh Ulhas
University of Dayton

Advisor: Dr. Vijayan K. Asari

Emotional state assessment of humans has been traditionally studied using various direct physiological measures and psychological self-reports. In this dissertation, we demonstrate that the variability in emotional and physiological states during stressful situations can be represented by analyzing indirect facial dynamics and thermo-grams. The two emotional states of challenged and threat responses in which a human under a stressful event can cope or not cope; respectively are investigated. This study uses several standoff non-invasive sensing technologies including 2D visible (VIS), Mid-Wave Infrared (MWIR) and a 3D Near Infrared (NIR) imaging cameras to examine the effect of thermal feature changes on the human face for both threat and challenge reactions. A
methodology for accurately extracting facial features from thermal (MWIR) sensors and classification of these signatures against psycho-physiological ground truth are presented in this dissertation research. A threat/challenge scenario for individuals using two different human study experiments namely: “false opinion study” and “false behavior study” have been created to elicit the emotional state changes. Classification studies from participants are conducted to understand the thermal facial feature changes with respect to the emotional responses. Ground truth of the threat/challenge emotional responses was conducted with direct physiological measurements and psychological self-reports which showed time-differentiated emotional response characteristics.

The proposed epoch-based windowing methodology extracts statistical features from thermal facial regions to understand the small temperature changes that happen over the time-windows (e.g., epochs) of different emotions. The feature vector for representing different emotional state epochs is defined by performing Eigen analysis on the statistical features of thermal sensor data. A K-Nearest Neighbor (kNN) classifier is used to discriminate the threat and challenge emotional states. An attempt to extract variations in local intensity distributions in the visual and NIR facial region to understand the muscle variations that happen over the course of threat and challenge emotions has also been studied in this research. Region-based feature sub-selection illustrates that certain regions (e.g., forehead and nose) of the face are more useful than other regions for threat/challenge detection. Experimental evaluations performed on a set of data collected in a previously mentioned “false behavior study” show that the proposed method
provides a unique and effective way to classify threat and challenge responses using thermal facial signatures. Research work is progressing to validate an optimal sensor suite outputs against established psycho-physiological variables indicative of human state responses to a stressful event.

The novelty of the research is an indirect method of detecting threat/challenge emotional responses from thermal facial features that could replace direct physiological measurements for quick and non-destructive evaluation of human behavior.
Dedicated to my wife, son and family.
ACKNOWLEDGEMENTS

I am grateful to Dr. Vijayan Asari, my dissertation advisor, for his support, guidance and his invaluable motivation. I am also grateful to the Office of Naval Research for the research grant (#N0014-10-1-0295), which provided the support needed.

I would like to thank other members of my dissertation committee: Dr. Russell Hardie, Dr. John Loomis and Dr. Erik Blasch for being a part of my committee and being very supportive. I would also like to thank Dr. Douglas Petkie and Dr. Julie Skipper from Wright State University, who kindly offered me help during the research work. In particular, I would like to acknowledge Dr. Tamera Schneider for providing us with the ground truth for the experiments. I would like to thank Rebecca R. Riffle who helped develop the protocol, created the surveys, and did schedule coordinating and running the subjects. I also wanted to thank Meena El-sheer, Matt Sherwood, Carl Cross, Gary Kash, James Trame, Jeremy Lewis, Natalie Gilbert, Adrian Johnson and Melissa Peterson for their efforts in data collection, image pre-processing and feature tracking. I appreciate all members of the Vision Lab at the University of Dayton for their support. I would like to specially thank my family who has continuously supported me in my effort.
TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................ iii
DEDICATION ...................................................................................................................................... vi
ACKNOWLEDGEMENTS ................................................................................................................ vii
LIST OF FIGURES .......................................................................................................................... xii
LIST OF TABLES ............................................................................................................................... xvi

CHAPTER I. INTRODUCTION ........................................................................................................ 1

1.1 Emotional state assessment via facial imaging ................................................................. 1
1.2 Psycho-physiological model ............................................................................................. 4
1.3 Sensors employed in the research ....................................................................................... 5
1.4 Methodology and technical approach with system architecture ...................................... 6
  1.4.1 Data collection ................................................................................................................ 8
  1.4.2 Preprocessing .................................................................................................................. 9
  1.4.3 Feature extraction ........................................................................................................... 10
  1.4.4 Classification .................................................................................................................. 11
1.5 Objective of the research (Problem statement) ................................................................. 13
1.6 Outline of the dissertation document .................................................................................. 14
### CHAPTER II. LITERATURE REVIEW

- **2.1** Detecting deception from remote sensing ......................................................... 15
- **2.2** Threat – challenge continuum ........................................................................... 16
- **2.3** Facial observations as an indicator of stress ..................................................... 17
- **2.4** Stand-off sensing technologies ......................................................................... 19
- **2.5** Summary of Related Research .......................................................................... 24

### CHAPTER III. BACKGROUND

- **3.1** Experimental design and set up ........................................................................ 27
  - **3.1.1** Study 1 (false opinion - Low stake deception) ............................................. 30
  - **3.1.2** Study 2 (False behavior - High stake deception) .......................................... 31
- **3.2** Data acquisition ................................................................................................ 32
- **3.3** Image pre-processing ........................................................................................ 33
- **3.4** Summary ........................................................................................................... 35

### CHAPTER IV. FEATURE EXTRACTION AND CLASSIFICATION

- **4.1** Feature extraction on thermal and visible data .................................................. 36
  - **4.1.1** Facial point tracker ....................................................................................... 37
  - **4.1.2** Automated face segmentation ....................................................................... 39
  - **4.1.3** Visible and thermal features ......................................................................... 40
- **4.2** 3-D Surface features ......................................................................................... 47
  - **4.2.1** 3-D surface data ............................................................................................ 47
5.7 Pattern analysis for different regions of the face with varying epochs .......... 87
5.7.1 Pattern analysis experiment I: whole face region + all the epochs .......... 88
5.7.2 Pattern analysis experiment II: whole face region + first two epochs .... 89
5.7.3 Pattern analysis experiment III: Forehead region + all epochs .......... 90
5.7.4 Pattern analysis experiment IV: Forehead region + first two epochs .... 91
5.7.5 Pattern analysis experiment V: Nose region + all epochs ................. 92
5.7.6 Pattern analysis experiment VI: Nose region + first two epochs ........ 93
5.7.7 Pattern analysis experiment VII: (Forehead + Nose) region + all epochs ................................................................................................................ 94
5.7.8 Pattern analysis experiment VIII: (Forehead + Nose) region + first two epochs ........................................................................................................ ……95
5.7.9 Experiments summary ........................................................................ 96
5.8 Pattern analysis for test data set with 30 participants for Study 2 .......... 97
5.9 Pattern analysis for test data set with 40 participants for Study 2 .......... 100
5.10 Deception based classification ................................................................. 102
5.11 Sample size versus total number of features ............................................ 104
5.12 Summary ............................................................................................... 108

CHAPTER VI. CONCLUSIONS AND FUTURE WORK ...................................... 109

REFERENCES .............................................................................................112
LIST OF FIGURES

Figure 1-1. Psycho-physiological model of human state [10, 11] ................................. 4
Figure 1-2. Pattern recognition architecture (adapted from [16]) ................................. 7
Figure 1-3. System architecture .................................................................................... 8
Figure 3-1. Experimental settings ................................................................................ 29
Figure 3-2. Test checkerboard patterns for image co-registration with fiducial points in both the VIS and MWIR ................................................................. 34
Figure 4-1. Facial point tracking .................................................................................. 38
Figure 4-2. Results of automated face segmentation can be reviewed in video format ......................................................................................................................... 39
Figure 4-3. Thermal signature for baseline and task ...................................................... 45
Figure 4-4. Feature visualization graphical user interface .............................................. 46
Figure 4-5. 3-D surface shape, 3-D normal, texture image overlay ............................... 48
Figure 4-6. (a) 3-D Face model obtained (b) depth map of face model (c) 3-D face imagery in frontal view with texture (d) principal curvature $k2$ (e) principal curvature $k1$ ......................................................................................................................... 49
Figure 4-7. Surface primitives of depth map ................................................................. 51
Figure 4-8. Surface primitives with 8 different features .......................................................... 52
Figure 4-9. Glabella energy tracking for threatened participants ....................................... 54
Figure 4-10. Glabella energy tracking for challenge participants ....................................... 55
Figure 4-11. Multi-modal feature framework ....................................................................... 56
Figure 4-12. Summary data for participants for study 1 and study 2 ................................. 57
Figure 4-13. Physiological ground truth .............................................................................. 59
Figure 4-14. Using appraisal ratio ground truth, participants are divided into two
        categories: threatened (z-score>0) and challenged (z-score<0). .......................... 60
Figure 4-15. PCA results of threatened (T) vs. challenged (C) participants ................. 64
Figure 4-16. PCA results of threatened + challenged (T+C) vs. neutral (N)
        participants ........................................................................................................... 64
Figure 4-17. Top-down approach to testing various classification schemes ................... 65
Figure 4-18. A block diagram of the multi-layer perceptron neural network ............. 66
Figure 4-19. Confusion matrix for the artificial neural network classifier using
        histogram feature. .................................................................................................. 67
Figure 5-1. Appraisal ratio for study 2 participants ............................................................. 73
Figure 5-2. Mean of top 10% hot pixel feature tracked over the baseline and task........ 75
Figure 5-3. Facial segments ............................................................................................... 76
Figure 5-4. Epoch based features ....................................................................................... 77
Figure 5-5. Time series based Eigen analysis ..................................................................... 78
Figure 5-6. Epoch based Eigen analysis ............................................................................ 79
Figure 5-7. Principal component vs. variance. ................................................................. 80
Figure 5-8. Biplot describing PC 1 vs. PC 2 for epoch based features......................... 81
Figure 5-9. Biplot describing PC 2 vs. PC 3 for epoch based features......................... 82
Figure 5-10. Loadings for each of the three principal components. ............................... 84
Figure 5-11. Scatter plots with PC 2 and PC 3 for forehead, nose and fusion of forehead and nose, classification accuracy using K-NN for forehead, nose and fusion are 70%, 80% and 85% respectively................................................................. 85
Figure 5-12. Face segmentation showing forehead (green), eye-periorbital (cyan), cheek (yellow), nose (red), mouth (blue) and neck (magenta). ........................................ 86
Figure 5-13. Accuracy estimation along with the confusion matrices for the whole face with all the epochs................................................................................................ 89
Figure 5-14. Accuracy estimation along with confusion matrices for the whole face with first two epochs................................................................................................ 90
Figure 5-15. Accuracy estimation along with confusion matrices for the forehead with all the epochs. ................................................................................................... 91
Figure 5-16. Accuracy estimation along with confusion matrices for the whole face with first two epochs................................................................................................ 92
Figure 5-17. Accuracy estimation along with confusion matrices for the nose with all the epochs........................................................................................................... 93
Figure 5-18. Accuracy estimation along with confusion matrices for the nose with first two epochs......................................................................................................... 94
Figure 5-19. Accuracy estimation along with confusion matrices for the fusion of nose and forehead for all epochs ................................................................. 95

Figure 5-20. Accuracy estimation along with confusion matrices for the fusion of nose and forehead for first two epochs ......................................................... 96

Figure 5-21. Data partitioning indicating additional (five lesser threat + five lesser challenge) test participants shown in the dotted rectangular box ......................... 98

Figure 5-22. Accuracy estimation along with confusion matrices for the fusion of nose and forehead for first two epochs for lesser threat and lesser challenge participants ........................................................................................................ 99

Figure 5-23. Data partitioning indicating additional (ten lesser threat + ten lesser challenge) test participants shown in the dotted rectangular box ......................... 100

Figure 5-24. Accuracy estimation along with confusion matrices for the fusion of nose and forehead for first two epochs for lesser threat and lesser challenge participants ........................................................................................................ 102

Figure 5-25. Deception based participants in threat, challenge and neutral category. ... 103

Figure 5-26. Deception participants scatter plots .......................................................... 104

Figure 5-27. Threatened feature data with positive slope........................................... 106

Figure 5-28. Challenged feature data with negative slope........................................ 107
LIST OF TABLES

Table 1-1. Applications using facial features to understand emotions ............................... 3

Table 3-1. Sensor platform describing different wavelength, pixel sizes and sensitivity. ............................................................................................................... 33

Table 5-1. Summary data for lie condition for study 1 and study 2 ................................. 71

Table 5-2. Region based classification. ............................................................................ 87

Table 5-3. Accuracy estimation for all experiments ......................................................... 97
CHAPTER I

INTRODUCTION

Emotional state assessment is a complex multidisciplinary field that involves recognizing different internal human emotions using psychological self-reporting and physiological monitoring. To empirically investigate emotions, clinicians and social psychologists use self-reports and other direct physiological measurements such as blood pressure, galvanic skin response, and electro-cardiogram (ECG). This dissertation research aims at determining an indirect method using thermal imaging for analyzing the facial dynamics and thermo-grams using advanced imaging and pattern recognition techniques to predict the human emotional state by examining a couple of minutes of data after elucidating a threat/challenge scenario.

1.1 Emotional state assessment via facial imaging

Analyzing a human face involves physical measures that include visual, acoustic and behavioral modalities [1]. Visual modalities include facial expression, head gesture, eye movements, gaze, etc. Acoustic modalities would include expression changes in voice such as variations in volume, pitch and semantic content. Behavioral modalities are
heavily dependent on the specific application. For example, traditional methods of behavior assessment include a cognitive task analysis in which moving a mouse determines a behavioral modality when a human subject is working on a computer. Likewise, eye-tracking has been used to determine where a subject is looking on the computer for decision making, workload, attention, and stress assessment. Beyond eye-tracking and visual facial recognition, this dissertation research uses the thermal modalities to extract features of facial changes due to changes in emotion. Identifying an emotion automatically with pattern recognition algorithms in real-time is advantageous in various fields such as stress determination in jobs where lives are at risk (e.g. air traffic controllers). Recent technological advances in imaging, computing, computer-vision and pattern recognition have made it possible to analyze the facial visual modalities very effectively [2]. Human emotions trigger specific facial activity as external signals. These external signals can be captured using non-contact and non-invasive type of sensors. There are a variety of applications which use facial features to understand the emotions; some of them are mentioned in the Table 1-1.
Table 1-1. Applications using facial features to understand emotions

<table>
<thead>
<tr>
<th>Field</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology</td>
<td>Interpreting the facial emotions of children with Autism Spectrum Disorders</td>
</tr>
<tr>
<td>Biometrics</td>
<td>Video surveillance applications</td>
</tr>
<tr>
<td>Law Enforcement</td>
<td>Lie detection</td>
</tr>
<tr>
<td>Fatigue detection</td>
<td>Monitoring a driver’s drowsiness by tracking the head pose and eyes</td>
</tr>
<tr>
<td>Operator performance</td>
<td>Cognitive workload monitoring</td>
</tr>
<tr>
<td>User experience research</td>
<td>Usability testing, gaming research, simulation and virtual reality</td>
</tr>
<tr>
<td>Consumer behavior research</td>
<td>Market research</td>
</tr>
</tbody>
</table>

While numerous articles relate to facial tracking in the above scenarios, three issues we explore in the dissertation have not been documented. The first contribution is that the purposes of facial feature response were not specifically determined from thermal imaging as related to determination of an emotional change. Data is collected from all the three modalities including visible, thermal and 3-D NIR cameras. The second is the combination of psychological ground truth data of emotional responses to support the facial features analysis. The third case is the use of epoch based feature responses to compress the number of features needed to distinguish an emotional response.

The next sections overview the psycho-physiological human state model along with the sensors and state measurements that provide measures of emotional reactions.
1.2 Psycho-physiological model

The model of the human state includes appraisals, emotions and associated physiological responses to a particular task, and yields aspects of a continuous state which ranges from a challenged to a threatened state [10, 11], as illustrated in the right column of Figure 1-1. The **challenged state** represents persons who believe they have good coping resources to deal with a demanding stressor and they generally display greater positive emotions and a greater cardiac output coupled with vasodilation. The **threatened state** reflects an individual who believes they lack sufficient resources to meet the demands of a stressor and thus they will experience greater negative emotions. A method to determine a threatened state is a cardiac output that has increased from a baseline resting state but is below that of the cardiac output of a challenged person.

![Psycho-physiological model of human state](image)

*Figure 1-1. Psycho-physiological model of human state [10, 11].*
The appraisals and emotions in this model of arousal regulation have shown reliable and valid measurements through self-reports, while the physiological measures are made with contact sensors that include impedance cardiography, electrocardiography (ECG), and blood pressure. Research has demonstrated that stressor appraisals (a person’s interpretation of an impending stressful situation) predict subsequent emotional, physiological, and behavioral performance patterns [12]. Both types of appraisal lead to physiological mobilization/reactivity. However, challenged participants have greater cardiac output coupled with vasodilation, systemically more blood is pumped out of the heart over time coupled with a more accepting vasculature [10, 11]. Threatened participants have moderately increased cardiac output, coupled with vasoconstriction, as more blood going into a more resistant vasculature.

1.3 Sensors employed in the research

Psychological research for the recognition of emotions from facial expressions has evolved over the years. Human emotions trigger specific facial expressions as external signals and other physiological changes as internal signals representing blood flow, heart rate, body temperature, etc. For most of the external signals, various non-contact sensors are used; mainly Electro-Optics (EO) sensors (e.g., visual and thermal) may be used to capture facial expression in video sequences [13]. As with the plethora of literature on visual facial recognition, these typically recognize a person and not their emotional response. Thermal imagery is dynamic and sensitive to the physiological processes in the body, with heat transfer from physiological processes affecting the surface temperature of
the skin. Thermal imagery could be collected for the analysis of human signatures to understand human internal states and emotions. Forward Looking Infrared (FLIR) systems, or ThermoVision cameras, have become the new standard for science and research in thermal imaging and measurement applications [14]. Geometric facial data from unique 3-D sensors could be used to provide significant improvements in facial expression and emotion recognition performance.

To measure physiological changes in the body during an emotion, several other sensors could be used. For example, a two lead Electrocardiogram (ECG) with electrodes was placed on the subject’s chest, Galvanic Skin Resistance (GSR) was measured on the chest band, and real time blood pressure measurement was measured in the non-dominant arm [15]. For the psychological changes, self-reporting (survey questionnaires) was employed where the participant answers questions before, during, and after the emotion eliciting video. The physiological and psychological data together would be used to estimate ground truths for a particular emotion. Next, we discuss sensor pattern recognition methods.

1.4 Methodology and technical approach with system architecture

The process of pattern recognition can be divided into four sequential stages: data pre-processing, dimensionality reduction, classification, and validation, as shown in Figure 1-2. As in most standard classification problems, a similar four-step process is followed: data collection, pre-processing, feature extraction and classification, shown in Figure 1-3. We leverage the same process for human state assessment and investigate the features
salient for emotional response classification. The steps are carried out independently for the psycho-physiological data that provides the ground truth, and the thermal sensor data under investigation. Ultimately, the aim is to assess the effectiveness of the thermal sensor data and epoch-based feature classification scheme in correctly identifying human emotional states. For both the psycho-physiological and stand-off sensor pathways, data is collected and processed. The data collected was separated into training and a testing set. The psycho-physiological data from the training set forms the ground truth to train the classifier over the extracted features. The trained classifier is then tested using the testing scenario to determine the verification of the proposed approach.

![Pattern recognition architecture](adapted_from_[16])
1.4.1 Data collection

**Physiological devices and Psychological tools**: The data acquisition process involves capturing data from the physiological instruments like blood pressure (BP) and Electrocardiogram (ECG) during the experiment. A self-reporting questionnaire is filled by each participant before the task (baseline), during the task and after the task.

**Non-contact imaging sensors**: It involves acquiring real time face data from three non-contact sensors: EO-camera manufactured by Basler (A202k), Thermal camera manufactured by FLIR systems (SC 6700) and a NIR spatial phase imaging 3D camera manufactured by Photon-X acquired real time facial videos of a person responded to a
threat/challenge scenario.

1.4.2 Preprocessing

The first pattern-recognition stage, data preprocessing, is critical as it determines the performance of all subsequent stages [17]. Data preprocessing techniques include removal of direct current (DC) offset, sensor auto-scaling and vector normalization [16].

Artifact removal: The self-report survey data and key experimental variables are entered manually by the subject under test. A frequency analysis is computed, which includes all variables (experimental variables, demographics, repeated administrations of surveys, pre-performance parameters, post-debriefing variables) to ensure the manual entered data are within acceptable ranges. Similarly frequencies for the physiological variables are investigated and validated for the emotional responses. Further descriptive statistics for the physiological variables to examine means, standard deviations, maximum and minimum values are computed. Generally, values should be within three standard deviations of the mean. If there are deviations from normal confidence interval (three standard deviations from the mean) then a participant’s pattern of data is examined to investigate whether the individual may be a physiological outlier. There are numerous ways to address those with values outside the range and decisions are recorded. The overall process is typically referred to as artifact removal.

Image registration: Image co-registration is performed using test pattern imagery collected pre-baseline for each subject. The test pattern provides contrast in both the visible and MWIR bands of the spectrum. Image registration (VIS/MWIR and VIS/NIR)
is performed using an affine transformation [18].

1.4.3 Feature extraction

The next stage in pattern recognition is feature extraction for dimensionality reduction. The goal of feature extraction is to reduce the dimensionality of the measurement space while preserving the information content. This is typically done using a linear transformation, i.e., by multiplying the measurement vector with a projection matrix. The two well-known feature extraction techniques are Principal Component Analysis (PCA) [19] and Linear Discriminant Analysis (LDA) [20]. PCA only finds projections of maximum variance, regardless of class labels and, therefore, is not guaranteed to find the directions of maximum discrimination. LDA, on the other hand, finds projections that maximize class separability, making it more appropriate for classification problems. There are also nonlinear transforms such as Sammon’s non-linear mapping [21] and Kohonen’s Self-Organizing Map [22] that can be used to perform dimensionality reduction. Recent methods include score fusion for facial recognition from EO and IR cameras using a combination of Face Pattern Byte (FPB), LDA, elastic bunch graph matching, and Hidden Markov Models (HMM) [23].

Psychometric reliability analysis and variable creation: Psychometric analysis for scales measuring primary appraisals, secondary appraisals, positive affect and negative affect are conducted. The appraisal ratio score for each of the subject is calculated as a ratio of the two variables (primary appraisal/secondary appraisal). The autonomic reactivity scores are created for physiological variables by subtracting the baseline from the task
values. These variables are computed for heart rate, cardiac output and mean arterial blood pressure. Vascular resistance is derived from heart rate and cardiac output. The baseline and task vascular resistance values are computed and then reactivity scores are created for this appraisal ratio score.

**Facial feature tracker:** Motion extraction techniques involve feature point tracking (pivot points), difference images, etc. where certain points on the face are tracked throughout the expression. The movement of the feature points can be measured by tracking the displacement between corresponding feature points. The feature pivot points can be initialized in the first frame either manually or automatically and then using a tracking algorithm (such as Lucas Kanade Tracker [24]), where all the feature points can be tracked over a series of frames for the entire expression period. Based on the pivot points, the face is divided into certain segments and statistical and textural features for visual and thermal image segments are computed.

1.4.4 Classification

The third stage in pattern recognition deals with classification, which can be performed with a number of techniques, including K-nearest neighbors (K-NN) [25] and Artificial Neural Networks [26]. These techniques can correctly predict the class label of a given unknown sample. A K-NN classifies an unknown sample by finding the closest samples in the database and assigning the unknown sample to the majority class among those neighbors. Artificial Neural Network (ANN) classifiers can model complex non-linear mappings and make weak assumptions about the probabilistic properties of the
underlying distributions of the data. An output of the classification step is the selection of parameter settings for the feature classification model. Testing the model against truth data determines the error rates of the final model by means of validation techniques.

This work uses PCA and a K-NN classifier for pattern recognition of the emotional states. PCA is a more robust method that is not subject to overfitting and is not tied to classification problems (as opposed to LDA). When the training data set is small, PCA may do better than LDA. Secondarily, the K-NN rule was selected as it can approximate the optimum Bayes error rate in the large sample case, yet works relatively well with sparse datasets.

**Ground truth estimation:** Broadly speaking, there are two different types of ground truth data available in this study for emotional state assessment. One is based on physiology and other is based on psychology. The physiological ground truth depends on vascular resistance variable, while the psychological ground truth depends on appraisal ratio variable. A higher value appraisal ratio denotes greater threat and a lower value denotes a challenge emotional state. As verified from the psycho-physiological analysis, the appraisal ratio provided the ground truth estimation for our classifier system.

The classification block helps to map the features to discrete expressions or action units (AUs). One of the spatial approaches is using neural nets (multi-layer perceptron). The neural nets are applied on either Principal Component Analysis (PCA) based features or just raw facial image data. Many times, just a simple K-Nearest Neighbor is enough to discriminate an emotional response if there is good separation between features. A spatio-
temporal approach such as a Hidden Markov models (HMM) [27] allows the classifier to model the temporal information of the emotion.

1.5 **Objective of the research (Problem statement)**

The primary objective of the proposed research effort is to develop an emotional state assessment system using multimodal non-contact sensors; namely visual, MWIR thermal, and Near-IR spatial phase imaging 3D sensors. In addition to the data collection, a state classifier based on specific features from the thermal sensor data is needed. The classifier can be validated using the psycho-physiological index as ground truth which was collected in the threat/challenge experiments. Eventually, a sensor model would be created that would establish an association between human facial signatures captured from a standoff sensor system and a separately validated psycho-physiological assessed state from an elicited emotional response.

The specific objectives of this dissertation research are the following:

1. Determine a non-contact method to discriminate threat and challenge human states.
2. Design of a feature extraction system based on thermal facial signatures.
3. Characterize threat and challenge participants based on deception and non-deception conditions.
4. Implement a classification strategy to separate threat and challenge participants.
5. Explore a thermal facial region based classification model that characterizes threat and challenge response states.
6. Perform experimental evaluations based on two separate subject studies (false opinion and false behavior) to validate the classification model.

1.6 Outline of the dissertation document

The rest of the dissertation is divided into specific chapters discussing various aspects in this research. Chapter 2 provides a brief literature review based on emotional state response assessment. Chapter 3 discusses the data acquisition, psycho-physiological setup, and the experimental design and data preprocessing. Chapter 4 presents the feature extraction and feature classification techniques. Chapter 5 focuses on results and discussion, and Chapter 6 presents conclusions and future work.
CHAPTER II

LITERATURE REVIEW

A brief survey of the relevant literature regarding emotional state assessment is presented in this chapter. The emotional state assessment is a complex multidisciplinary field with theoretical roots in psychophysiology. To understand different emotional states, researchers in psychology utilize different assessments including self-reports, physiological monitoring and observation. Self-reports assess emotional experience; physiology assesses blood flow and sympathetic nervous system, and behaviors indicate approaching or avoiding a stimulus. This dissertation research aims to analyze the facial dynamics and thermo-grams during a high-stakes deception situation, using psychophysiological metrics as well as advanced imaging and pattern recognition techniques to remotely predict the human emotional state.

2.1 Detecting deception from remote sensing

Detecting deception using non-verbal cues (eye-contact, nervous laughter, and hurried speech) is a complex task and the classification rate achieved was around 50 percent [28].
Burgoon [28] explains that deceivers were more uncertain and vague, as well as showed more negative affect. To our knowledge, previous research has not found any prototypical facial expressions for deception. Using current technology in remote sensing, we hypothesize that we can detect facial expression changes, facial action units, or changes in face temperature.

Another interesting study reveals that if a person was concerned that their deception was likely to be detected, it would in turn cause a change in his brain activity and would probably increase his anxiety [29]. This study was evaluated using functional MRI (FMRI) and suggested that belief in lie-detection efficacy modulates a subset of the processes involved in producing deception. Some deceivers were not thinking about being caught and would manage to provide low emotional reactions, however the innocent ones end up in higher levels of anxiety resulting from stress.

2.2 Threat – challenge continuum

The stress process begins with a person’s appraisal of an event and every individual appraise situations differently, causing their stress response to differ [10, 11]. In early 1966, Lazarus mentioned that there are three kinds of stress: harm, threat, and challenge [30]. Harm refers to psychological damage that has already been caused. Threat is anticipation of future harm implications that has not yet occurred and is associated with negative emotions. Threat appraisals result when stressor responses outweigh coping resources [11]. Challenge results from difficult situation that people are confident of overcoming and is always related to positive emotions. Challenge appraisals occur when
a stressor is commensurate with coping resources. T. R. Schneider has demonstrated how neuroticism influences both psychological and physiological stress responses [10].

The above research indicates that threat and challenge responses are human emotional states which are distinctive. Research shows that challenged participants experience more positive and less negative effect than threatened participants [10, 11]. These two groups have also been distinguished by their physiological pattern differences. During a stressor, challenged participants have increased cardiac output into a more accepting vasculature, whereas threatened participants have somewhat more blood pumped into a more resistant, constricted vasculature [10-12]. High-stakes situations are stressors that engage stress responses in participants. Research indicates that threat and challenge appraisals are appropriate for discerning the human state and emotional activity, however, little research has been shown to measure these states remotely.

2.3 Facial observations as an indicator of stress

The human state and emotional assessment can be defined as a participants’ stress response to a specific stimuli that includes behavioral, physiological and visual components. Behavioral measurements in stress studies are subjective, physiological measurements are more impartial. However, visual facial observations also provide quality indicators of stress [31].

Visual modalities include facial expression, head gesture, eye movements, and gaze. The domain of human state assessment and emotion recognition from a psychological perspective has been expanding over several years. Ekman and Friezen
[32] have shown that facial expressions may be linked to specific (i.e., basic) emotions, and that these basic emotions are relatively universal across human cultures. Stemming from his work with Tomkins [33], Ekman has developed a systematic method of categorizing human facial expressions known as the Facial Action Coding System (FACS). The FACS system is based on facial muscles movements and it is defined in terms of action units (AUs). The FACS system can be used to manually code any anatomically possible facial expression. This technique is extremely useful to reveal novel facial movements from the landmark points. More than 7,000 different combinations of action units have been recognized in different facial expressions. One of the major disadvantages to FACS is it is extremely laborious and the FACS coder has to go frame by frame to detect changes in AUs which is time consuming.

Another widely used approach uses facial electromyography (EMG), to try to understand facial movements. The EMG process involves attaching electrodes on the skin of the face to measure the intensity of muscular contractions. The EMG technique was not determined not to be a viable option while collecting our research participant data for number of reasons: 1) EMG majorly focuses on muscle activity and eventually to action units which would not be widely adopted, 2) EMG would not qualify as a remote sensing or standoff type sensor with minimal intrusion, and 3) lastly, EMG would make the participants uncomfortable with electrodes sticking on their face when they perform the stressor task which complicates the measurement of threat/challenge responses.
2.4 Stand-off sensing technologies

Human emotions are thought to trigger specific facial activity as external signals (although facial activity is clearly not the single source of emotional expression). We hypothesize that these external signals can be captured using non-contact standoff sensors. The performance of an emotional state assessment depends heavily on the nature of the data. The data can be classified into six categories [34] as Two-dimensional (2-D) – dynamic data sequences, thermal-static data, thermal-dynamic data sequences, three-dimensional (3-D)– static data and three dimensional (3-D)-dynamic data sequences.

Several approaches have been researched to classify human affective states using facial expressions. One of the most common types of non-contact sensor is the traditional visual sensor system which can capture 2-D static images as well as 2-D dynamic video sequences. Yacoob et al. has proposed a mid-level symbolic representation for the spatial and temporal data using linguistic and psychological considerations [35]. The algorithms make use of optical flow to identify the motions caused by human facial expression. The system achieved a recognition rate of 86% for smile, 94% for surprise, 92% for anger, 85% for fear, 80% for sadness, and 92% for disgust for a total of 105 emotions examples. This method is particularly favored for simple expressions. A lot of participants had to express emotions which were artificially induced (such as fear and sadness which are hard to induce). Essa and Pentland have developed an advanced computer vision system to probabilistically characterize facial motion and muscle activation, thus developing a
new and more accurate representation of human facial expressions termed as FACS+ [36]. They modeled the face using optical flow method along with geometric, physical and motion-based dynamic models. Their expression recognition accuracy is close to 98% on a database of 52 sequences. Y. Tian et al. [37] recognized the changes in the action units (AUs) of the FACS using conventional 2-D images. Tians’ method makes use of geometric and appearance based features to extract information from static and dynamic image sequences. Geometric features are a set of points which can represent different facial components. Appearance-based features normally describe the texture changes during an expression.

J. Bailenson [38] et al. performed real time classifications of emotions (sadness or amusement) using facial feature tracking and physiological responses and recorded a classification rate of nearly 95%. One of the unique things is that their paper analyzes physiological features and uses the physiological features along with facial features to recognize emotions. In some cases, it has been shown that the physiological levels in detection of sadness outperform facial expression data analysis. The paper uses trained psychological coders labels for participants’ emotions. The ground truth reliability is heavily dependent on the inter-coder reliability to correctly labeled emotions. Most of the visible cameras data help perform facial feature point tracking that could be used to analyze emotions. This may not be enough to understand complicated emotions such as challenge/threat responses, so there is a need to explore other sensors like thermal sensors.
Thermal imaging or Mid-Wave Infrared (MWIR) imaging has gained a lot of popularity in several fields such as Chemical, Biological, Radiological, Nuclear and Explosives (CBRNE) detection, surveillance systems, security, military and health industry. Another advantage of thermal imaging is that it does not depend on illumination variations as compared to visible EO sensor system. Thermal imaging is a method to explore facial changes in temperature during the state assessment. O’Kane et al. [39] have demonstrated noticeable changes in the thermal signature of the human face during breathing, muscle tension, aerobic exercise, and during aggressive playing of video games. These results suggest that thermal imaging is a promising technology that could be used to gain insight on the perception of human state assessment and also to understand underlying internal states.

I. Pavlidis et al. [40] have shown evidence of a unique way to capture high definition thermal images of the face for detecting deceit. Exploring thermal images and detecting deceit has accuracy comparable to the polygraph examination. To attach electrodes and to perform a security screening at the airport using a polygraph mechanism for each individual person is almost impossible because of the amount of time needed. Using thermal imaging of a face gives a specific thermal signature for different emotions. In the paper [40], an experiment is conducted with twenty participants and they were asked to stab a mannequin, rob it for $20, and then prove that they are innocent. The thermal imaging was successful in correctly classifying 6 out of 8 participants who were guilty. Using thermal imagery as a stand-off sensor in turn helps to
measure and analyze psychological responses without contact sensors. Another study [41] measures the startling effect using thermal imaging. Facial thermal signatures changes have been seen near the periorbital and cheek regions for subjects after fright eliciting experiments. The study in [41] shows that thermal signatures of the face help us to determine the psychological state of a person. However, the above mentioned studies did not provide any pattern recognition analysis of thermal signatures to classify the emotions. Next, we discuss studies which provide a detailed pattern analysis of the respective thermal signatures.

Using the facial temperature difference images and using binary pattern from specific facial regions were assessed using neural nets for classification. Yoshitomi et al. [42] could distinguish happy, surprised and sad expressions with a recognition accuracy of 90%. They used 17 normalized facial area regions for characterizing each emotion. Liu and Wang [43] analyzed a facial temperature sequence data and computed statistical features and temperature difference histogram features. Further, Hidden Markov Models (HMM) were used to discriminate happiness, disgust and fear with a recognition rate of 68.11%, 57.14% and 52.30%; respectively. The results also demonstrated the temperature information of the forehead is more useful than other regions of the face. They used samples from the USTC-NVIE (natural visible and infrared facial expression) database to evaluate their results [44]. All the research demonstrate that, thermal cameras could be used as a non-contact, non-invasive way to detect the changes in the temperature across the face.
Jarlier et al. [45] show that the thermal changes of the face are caused by the changes in the facial muscle contractions. The FACS coders are trained to produce different action unit combination at various intensities. These changes in action unit combination eventually cause the thermal patterns which can be classified using a PCA decomposition of the thermal signal. One of the things to be noted is that all the coders are forced to certain emotions; which makes it difficult to detect and characterize spontaneous expressions.

In recent years, 3D range data has been used for face recognition, face expression analysis and emotion recognition. One of the major advantages to using 3D thermal range data is that it is pose and illumination invariant [46]. Just like 2D data, 3D data can be used in both the static and dynamic image space. There are several ways to acquire 3D data; one of the most popular and earliest technologies is to use two stereo cameras which will create a continuous point cloud. The State University of New York has developed 3D facial expression recognition (FER) database which uses two stereo cameras for 3D capture and another video camera for overlaying the texture [47]. The FER database is publicly available and used by the FER community [47]. Most of the data collected for the 3D Face Recognition Grand Challenge database is acquired using laser scanning [48]. A newer patented technology, termed 3-D spatial phase imaging, uses a single camera lens, and created a high resolution real-time 3-D point cloud without scanning or structured lighting [49]. This technology was used to extract features like area, volume and height of a surface and could be used to identify micro-expressions and eventually
could be helpful in human state assessment.

2.5 Summary of Related Research

The motivation of the dissertation research is to design a sensor model based on a stand-off sensor to detect the changes in face during deception in the dissertation research. Current methods exist for facial recognition primarily focused on visual sensors, but current methods are emerging for thermal sensors. Thermal sensors could be designed to detect stressor conditions; however, no research directly connects thermal facial imaging to detection of stressor responses of threat and challenge. The stress reactivity varies under deception conditions, because the act of deceiving is considered a stressor itself. The validity of remote sensing of facial activity is achieved by examining the stress responses with respect to appraisals and physiology.

A summary of different stand-off technologies capable of remotely sensing is discussed. Among all the three technologies: Visible, Thermal and 3-D, there has been no literature that we found that uses any of the three modalities to discriminate threat and challenge data from which to determine which features to collect and process. In the literature review, there were no papers quantifying threat and challenge in terms of action units. Therefore, using action units to quantify the responses requires investigation. Note: facial point tracking, could be used; but through investigation, we found that it did not provide enough information to distinguish threat and challenge participants.

The literature pertaining visible camera and 3-D camera emotion recognition are
focused on detecting action unit activation. At this point, there has been no FACS coding established for threat and challenge data, so an alternative to distinguish our emotions is investigated. If there was a way to characterize threat and challenge using FACS, then the visible camera and 3-D feature point tracking would determine the action unit activation and eventually classify the emotion.

Most of the literature we cited above relates to thermal imaging as it was deemed a needed area of exploration. Three key papers that are most closely related to the dissertation research include thermal imaging emotion detection [45], deception experiments to elicit stressor responses [40], and facial feature detection [43]. Thermal imaging has been successful in quantifying an emotion and also been used to determine the intensity of the emotion [45]. Therefore, the dissertation research focuses on the thermal imaging and a pattern analysis of thermal features for successful detection of stressor responses. The dissertation research focuses on thermal imagery to detect behavioral indicators of changes in emotion during a high-stakes situation which, to our knowledge, is the first analysis in this area. I. Pavlidis et al. [40] research was a great motivation for the dissertation, because it motivates us to detect threat – challenge emotions in the context of deception and non-deception based participants. The dissertation Study 2 was similar to the research paper study where they steal $20 and try to prove that they are innocent [40]. Certain regions of the face seem to provide information necessary for the classification as denoted in the paper [43]. The dissertation research aims at eliciting true spontaneous responses from the participants during the
experiment instead of forced response as described in the paper above by Liu and Wang.

A frame by frame analysis of 3-D mesh becomes extremely time consuming and memory intensive. The iterative closest point algorithm is used to find the differences in the mesh. The use of principal curvatures and other shape statistics to infer knowledge about the structure of a surface at a given point [50] has been employed. The extracted surface features are simplistic for fast processing, yet representative of the shape of a given surface. The process of reconstructing the 3-D data and analyzing the 3-D surface primitives is still being pursued as 2D thermal facial features were found significant to detect emotional responses. We note that 3D imaging was explored in the case that the subject was directly looking at the camera, but finding the robust features for various viewpoints would require more subject data collection.
CHAPTER III

BACKGROUND

This research was part of a collaborative project between Wright State Research Institute (WSRI) and University of Dayton Research Institute (UDRI) supported by Office of Naval Research with research grant (#N00014-10-1-0295). WSRI designed the psychological protocol for elicitation of threat and challenge data. The raw data with some initial preprocessing was made available for performing the automatic feature extraction and classification aspects in the project. All the data was collected at the WSRI facility. Details of experimental set up, overall psychological analysis and data acquisition procedures are presented in this chapter.

3.1 Experimental design and set up

The participants were asked to sit in a comfortable chair and various physiological sensors were measured as shown in Figure 3-1. The heart rate and blood flow were measured using sensors comfortably attached to the skin, around the torso and neck. Blood pressure cuffs were wrapped around the upper arm, and around the wrist for the
Cardiovascular signals such as ECG (electrocardiography) and ZKG (impedance cardiograph) were obtained with the Mind Ware Technologies instrument. Baseline impedance ($Z_0$) and the rate of change in impedance on a given heartbeat ($dZ/dt$) were used to derive measures of cardiac performance. These signals along with the ECG were used to estimate stroke volume, cardiac output, and contractile force, all of which determine cardiac reactivity. Cardiac output (CO) is the amount of blood pumped out of the heart on each heartbeat. CO is derived by multiplying heart rate by stroke volume. Cardiac output is combined with mean arterial blood pressure (MAP) ($CO \times 80 / MAP$) to estimate vascular resistance and determine vascular reactivity. To noninvasively assess impedance, two outer electrodes (one at the top of the neck and one several inches below the sternum) provide an alternating current across the thoracic cylinder and two inner electrodes (one at the base of the neck and one just below the sternum) measure the resistance to the alternating current. Blood pressure (SBP, DBP) and pulse rate (HR) were also continuously monitored (i.e., Mind Ware Technologies, Inc.). The blood pressure assessment was made on the non-dominant arm. Two experiments were conducted: Study 1 was False Opinion and Study 2 was High Stake Deception. For study 1, two assessments during the baseline and during the task were obtained. For study 2, a continuous blood pressure assessment enhanced reliability of the physiological measurements. For both studies, a 3D thermal imaging camera was in view of the participant stressor responses as shown in Figure 3-1.
Mean values of heart rate, cardiac output, and vascular resistance were calculated for the last minute of baseline to indicate resting rate and the first minute of each task to determine peak reactivity. The primary indexes for human state assessment from these measurements are cardiac output and vascular resistance (VR). These CO and VR measures were transformed into a single variable so as to yield a single physiology index. The *physiological index* is higher for a challenged individual than for a threatened one. The facial changes were measured using cameras (Visible, MWIR, NIR) during the study. The psychological responses were measured using pen and paper surveys.

A total 100 participants comprising of two subject studies is evaluated. Some of the data collected was not usable due to various reasons from which outliers were
removed. The first study (false opinion) included 44 participants (41 valid data sets) and second study (false behavior) included 50 participants (48 valid data sets).

3.1.1 Study 1 (false opinion - Low stake deception)

This section describes the experimental protocol for Study 1 in detail. The baseline measures for the physiology were recorded for 5 minutes and the last 10 seconds of the baseline measure was also recorded by the cameras. Then the participants are administered a baseline self-report emotion survey, demographics survey and an opinion survey for abortion and gay marriage attributes. The experimenter or researcher determines what is the participant’s strongest opinion and the participants are randomized to speak either towards or against the opinion with approximately \( n=22 \) participants in each category. The participants were either told they have to speak in favor of their opinion, or against their opinion, and that if the researcher believes that they are truthful, they will receive an additional bonus dollar amount, otherwise they will relinquish one-half of their original participant pay (high stake scenario). Immediately after these instructions, self-reported appraisals and emotions are assessed. Participants were provided two minutes to prepare the speech and two minutes to present a speech. After the two minute preparation time, participants are reminded to either speak to or against their opinion (depending upon random assignment) and this marks the beginning of the two minute speech task. The participant’s physiology is measured throughout the preparation and speech delivery. The sensor suite including (Visible, MWIR, NIR cameras) are capturing videos of the two minute speech delivery task.
Post–task assessment of appraisal and emotion are measured and participants are queried about their beliefs of how effective they were in persuading the interrogator. All participants were told that the judges were mixed in discerning their truthfulness, so each participant did receive full amount (original pay amount + bonus amount). Participants were also told they should not share the study details with others as it could affect the results of the study, and that others might not be judged as well and might forfeit their participation money, ending up with nothing (i.e., no monetary value for their efforts). A total of 44 participants in study 1 participated, 41 had usable data, of which 21 were in truth condition (spoke in favor of their opinion) and 20 were in false opinion condition (spoke against their opinion).

3.1.2 Study 2 (False behavior - High stake deception)

The overall approach and the protocol for Study 2 are almost similar to Study 1 however; there are few changes that we describe in detail. The baseline measures for physiology, psychological self-report and the multimodal imaging remains exactly similar to Study 1. The experimenter gives the participant an envelope that either contains or does not contain money. The participant is to either take the money that is there, or act like they are taking the money when there is none (rattle the envelope), so that experimenters are blind to the condition. The participant is told that their task is to deliver a speech about not taking the money and convene the three judges that they have not taken it. If the three researchers deem that the participant is truthful, they will receive an additional $40 bonus, otherwise they have to relinquish their $10 participant pay. Immediately after
these instructions, self-reported appraisals and emotions are assessed. Participants are
given one minute to prepare and two minutes to present a speech. After the 1 minute
preparation time, participants are reminded to convince the judge that they did not take
the money and the 2 minute speech task begins. The subject’s physiology is measured
throughout the preparation and speech delivery. Post task assessment of appraisal and
emotions remains similar to Study 1. All participants were told that judges were mixed in
discerning their truthfulness, so each participant was given the bonus money. Participants
were given the exact instructions after the experiment as in case of study 1 to protect the
nature of the experiment and not bias future participants. Of the 50 participants enrolled
in the second study, 48 had usable data. Of these, 25 spoke to their behavior condition
(about not taking the money when they did not take it), and 23 spoke against their
behavior condition (about not taking the money when they did take it).

3.2 Data acquisition

The imaging platform consists of three state-of-the-art systems presented in the Table 3-1
that include: visual (VIS), thermal midwave infrared (MWIR), and Near Infra-Red 3D
(NIR) camera. The images are collected at a distance of approximately two meters and
provide high resolution images for feature extraction. All cameras capture images at 30
frames per second, with the visual and thermal camera acquiring in time-synced mode
while the NIR data are collected asynchronously. The MWIR (FLIR SC 6700) camera is
calibrated with a blackbody and the ambient temperature and humidity are monitored
throughout the experimental period. The characteristics of the sensors are described in
Table 3-1. Sensor platform describing different wavelength, pixel sizes and sensitivity.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FLIR Systems SC6700</td>
<td>MWIR</td>
<td>3.0-5.0</td>
<td>640 x 512</td>
<td>15</td>
<td>14</td>
<td>125</td>
</tr>
<tr>
<td>Photon-X HD3D SPI</td>
<td>NIR</td>
<td>0.8-1.0</td>
<td>2048 x 2048</td>
<td>~7</td>
<td>12</td>
<td>30</td>
</tr>
<tr>
<td>Basler A202k</td>
<td>VIS</td>
<td>0.35-0.88</td>
<td>1004 x 1004</td>
<td>7.4</td>
<td>10</td>
<td>48</td>
</tr>
</tbody>
</table>

During the experiments, image data are collected during the last 10 seconds of the baseline epoch and during the entirety of each task, with the exception of Photon-X NIR 3D imagery, which is only collected for the first 30 seconds of each task due to the large data storage requirements.

3.3 Image pre-processing

The VIS and MWIR images were co-registered using test pattern images (that are collected pre-baseline) as inputs to our mutual information based registration algorithm that we tailored for thermal imaging based on coarse and fine region registration to improve accuracy. Mutual image registration is useful for surveillance and biomedical
imaging [51] and there are many methods described for mutual image fusion as compared to other multi-modal image fusion [52]. For a recent review of mutual image registration, see [18]. The test pattern provides contrast in both the visible and MWIR bands of the spectrum, as shown in Figure 3-2. Image registration (VIS/MWIR and VIS/NIR) is performed using our own mutual information-based algorithm that provides for both coarse and fine registration. Since the object (i.e., human face) depth is not large, an assumption of affine geometry is considered and thus, requires parameters for translation, rotation and scale differences.

![Figure 3-2. Test checkerboard patterns for image co-registration with fiducial points in both the VIS and MWIR.](image)

Operators, or experiment collectors, select 10-15 fiducial points in each of the test pattern images. Note that one-to-one correspondence is not needed during the selection process as the algorithm will examine all combinations of available points to find the set of analogous pairs. Six transform coefficients are returned, and then each MWIR or NIR
image is transformed to match the analogous VIS image using bilinear interpolation. As every pixel in the MWIR image (360k pixels) or NIR image (42M pixels) is updated, this routine must be efficient for real-time applications. Although an initial algorithm testing is performed on graphics processing unit (GPU)-accelerated machines, it was established that an optimized MATLAB code outperformed the GPU version. Currently, the registration process requires about 90 ms per image following manual control point selection.

3.4 Summary

This chapter presented the stressor design experiments and the overall experimental setup for acquiring data. It also described the development of the protocol for both the studies and explains the studies in detail. Initial co-registration of images for the MWIR and NIR was performed at the WSRI and normalized data was made available for feature extraction and classification.
CHAPTER IV

FEATURE EXTRACTION AND CLASSIFICATION

In this study, the key contribution is in the analysis of the thermal imaging to detect stressor responses of which a feature-based classifier has yet to be reported for accurately characterize stressor responses. The pre-processed and normalized data first undergoes a feature extraction procedure. The feature extractor includes facial point initializer and facial point tracker. These facial points are tracked on visible image and then corresponding points on thermal image are obtained, as we registered the data. The segmentation of the face is achieved using the tracked and derived points. The feature extractor determines the regions from individual segments of the face for both visible and thermal data. This chapter presents the feature extraction methods for all the three modalities. The method of estimating ground truth from appraisal data is also explained in detail. Once ground truth is obtained from the physiological and paper surveys, different classification strategies are applied to discriminate threat and challenge data.

4.1 Feature extraction on thermal and visible data

Commercial software for facial point tracking is employed. Note that the standard head
tracking products are not sufficient for the face segmentation analysis. The Visage-T Tracker (Visage Technologies AB, Linköping, Sweden) [53] is utilized, which returns 84 fiducial points based on the MPEG-4 Face and Body International Standard that is used for character and face animation. The software estimates the coordinates of point positions where no anatomical landmarks are present (e.g., on the cheeks) or when the point is out-of-frame, and not every point can be tracked in every frame (e.g., points inside the mouth). Among the set of returned control points, 49 of the points are retained and an additional 25 points are derived which would be useful for the segmentation scheme. These 74 points are subsequently used to divide the face into 29 non-overlapping facial segments as shown in Figure 4-1.

4.1.1 Facial point tracker

Although the Visage tracker features a real-time tracking mode that is a highly desired feature for an automatic system, this functionality was explored but is not reported in this dissertation research as we designed our own based on the needs for analyzing the data for the specific study. The image data is saved and the analyses are performed offline. Videos of the visible spectrum imagery (baseline and task) are created to be run on our tracker. Next, the operator makes a subject-specific mask on a ‘good’ baseline frame by scaling, rotating and dragging the vertices of a predefined mesh to match the current subject’s facial features; a trained operator can complete this process in 5-15 minutes.
Once the mesh is correctly overlaid, the tracking process is initiated and each frame of the video is processed automatically at a speed of 20-30 frames per second. The software is customized to return a file containing the following variables: frame name, tracking success (tracked/not tracked), $x$ and $y$ coordinates of each of the 84 points, and yaw, pitch and roll values for each successfully tracked frame. In general, the Visage product performs well, although some subjects and some frames cannot be tracked for reasons do to various operating conditions. Approximately 90% of the participant videos are tracked with few errors, and the tracking was accurate enough from our design to detect threat/challenge responses.
4.1.2 Automated face segmentation

Over all good performance image registration images, facial feature tracking resulted in robust facial segmentation. Figure 4-2 shows a representative visible and thermal frame onto which face segment boundaries are overlaid and segment numbers are displayed. For each collected frame, the tracking results are displayed at the top of the image (‘GOOD FRAME’, ‘NO VISAGE OUTPUT’, or a manually-determined ‘BAD FRAME’). Along the bottom of the frame, a display (left to right) indicating subject #, task # (1=baseline, 2=task), frame #, pitch, yaw and roll values is provided.

Figure 4-2. Results of automated face segmentation can be reviewed in video format.

Each frame is labeled with the result of Visage tracking ("GOOD FRAME" or "NO VISAGE OUTPUT"); a manual method of identifying a frame as a “BAD” frame is
implemented if poor tracking occurs. NO VISAGE OUTPUT and BAD frames are excluded from feature analysis. Segmentation is generally robust to head motion and to scale changes. Predictably, tracking is lost when pitch ("yes" motion) or yaw ("no" motion) is too great or when subjects obscure part of their face with their hand (e.g., coughing, wiping eyes, etc.). Tracking generally resumes within a few frames once proper position is resumed or the obscuration is removed.

4.1.3 Visible and thermal features

Features of interest in the visible imagery include wrinkling of the forehead, between the brows (glabella) or on the cheeks. These facial features are defined via a visual energy measure (based on the gray-level co-occurrence matrix) of these regions. These features are represented as forehead energy, energy in cheeks, and Glabellas’ energy. Eye gaze, blink rate and saccade velocity are extracted using other routines centered on iris detection.

From the thermal imagery, different kinds of features are extracted. Several statistical features (mean, minimum, maximum, standard deviation of pixel intensities) are calculated, as well as the mean of the top 10% hottest pixels in a segment. Through a thorough investigation, the mean of top 10% hottest pixels feature appears to be the most promising for classification. An implementation of a histogram of thermal feature is implemented, wherein each thermal image is segmented into a regular $4 \times 4$ grid, the histogram of pixel intensities in each grid segment is calculated and the cumulative distribution function for these data is formed. The segmentation of the face resulted in 16
separate regions with each region containing a specific part of the face. A histogram of intensity values are computed independently within each of the 16 regions. The resulting 16 histograms are combined yielding a (16×5=80) dimensional feature vector.

The five primary texture features are those defined in the thermal images as the Tamura features [54] that are useful in describing observed patterns in the imagery. Coarseness is the most fundamental feature in analyzing a texture, and refers to the size and number of texture primitives. A coarse texture contains a small number of large primitives (rocks), while a fine texture contains a large number of small primitives (sand). Contrast describes difference in intensity between neighboring pixels; a large intensity difference denotes high contrast. Roughness refers to tactile variations of a physical surface. Derived from coarseness and contrast, a rough texture contains angular primitives, whereas a smooth texture contains rounded primitives. Directionality describes the orientation of primitives. A directional texture has one or more recognizable orientation of primitives. And, regularity refers to variations of the texture primitive placement, e.g., a chessboard is composed of similar primitives which are regularly arranged. Finally, a method is implemented to extract information related to pores (pore count, pore area, pore activation, etc.) in the upper lip, brow/forehead, nose and chin segments.

Coarseness, contrast and directionality seemed important because these correlate strongly with human perception. The calculation of coarseness and contrast is explained in [54, 55]. The coarseness measure is calculated as follows.
For every image pixel co-ordinate \((m, n)\) in an image, the average at that point over the neighborhoods of size \((2^k \times 2^k)\) is defined in Equation (4.1) below

\[
A_k(m, n) = \sum_{i=m-2^k}^{m+2^k-1} \sum_{j=n-2^k}^{n+2^k-1} f(i, j) \bigg/ 2^{2k}
\]  

(4.1)

where \(f(i, j)\) is the gray-level at \((i, j)\).

Then at each point the differences are calculated between pairs of averages corresponding to the non-overlapping neighborhoods on opposite sides of the point in both horizontal and vertical directions as shown in Equation (4.2) and Equation (4.3) respectively.

\[
E^h_k(m, n) = \left| A_k(m + 2^{k-1}, n) - A_k(m - 2^{k-1}, n) \right|  
\]  

(4.2)

\[
E^v_k(m, n) = \left| A_k(m, n + 2^{k-1}) - A_k(m, n - 2^{k-1}) \right|  
\]  

(4.3)

At each point \((m, n)\), select the optimum size of the neighborhood \(K_{opt}\), which gives the highest difference value \(S_{opt}\) as shown in Equation (4.4) to Equation (4.7). The optimum average \(A_{opt}(m, n)\) is the value selected at an optimum size of the neighborhood \(K_{opt}\) and is denoted below in Equation (4.4)

\[
A_{opt}(m, n) = \sum_{i=m-2^{k_{opt}-1}}^{m+2^{k_{opt}-1}} \sum_{j=n-2^{k_{opt}-1}}^{n+2^{k_{opt}-1}} f(i, j) \bigg/ 2^{2k_{opt}}
\]  

(4.4)

The averages corresponding to the non-overlapping neighborhoods on opposite sides of the point in horizontal direction is denoted by \(E^h_{K_{opt}}(m, n)\) in Equation (4.5)
\[ E^h_{k_{opt}}(m,n) = \left| A_{k_{opt}}(m + 2^{k_{opt}-1}, n) - A_{k}(m - 2^{k_{opt}-1}, n) \right| \] (4.5)

The averages corresponding to the non-overlapping neighborhoods on opposite sides of the point in vertical direction is denoted by \( E^v_{k_{opt}}(m,n) \) in Equation (4.6)

\[ E^v_{k_{opt}}(m,n) = \left| A_{k}(m, n + 2^{k_{opt}-1}) - A_{k}(m,n - 2^{k_{opt}-1}) \right| \] (4.6)

The computation of maximum of \( E^h_{k_{opt}}(m,n) \) and \( E^v_{k_{opt}}(m,n) \) yields \( S_{opt} \) as denoted in Equation (4.7)

\[ S_{opt} = \max \{ E^h_{k_{opt}}(m,n), E^v_{k_{opt}}(m,n) \} \] (4.7)

Finally, take the average of \( S_{opt}(m,n) \) over the image is a coarseness measure \( F_{crs} \) for the image as in Equation (4.8)

\[ F_{crs} = \frac{1}{M \times N} \sum_{i=1}^{i_m} \sum_{j=1}^{i_m} S_{opt}(i,j) \] (4.8)

where \( M \times N \) denote the width and height of the image respectively.

**Contrast** \( F_{con} \) of an image is calculated by Equation (4.9)

\[ F_{con} = \frac{\sigma}{(\alpha_4)^{\frac{1}{r}}} \] (4.9)

The kurtosis \( \alpha_4 \) is defined by Equation (4.10)

\[ \alpha_4 = \frac{\mu_4}{\sigma^4} \] (4.10)

where \( \mu_4 \) is the fourth moment about the mean and \( \sigma^2 \) is the variance. The value of \( r \) is
calculated empirically by running different experiments with values of $r = 8, 4, 2, 1, \frac{1}{2}, \frac{1}{4}$ and $1/8$. Tamura [54] empirically found that $r = 1/4$ gave the best results.

A number of features of interest are defined, where the "feature" is the baseline-subtracted value for the task. One novel contribution of the feature analysis in our research was the use of epochs (spatial features extracted from the face over an interval of time). One reason for the epoch-based analysis was to match the number of features as related to the amount of data to validate the results. Using all features would require enormous amount of subject testing; however a small feature set can be verified against the amount of data collected from the two studies. Our time epochs of interest, including last three seconds of baseline, first three seconds of task, six-second sliding windows over the first 30 seconds of task, etc. as shown in Figure 4-3. The top part of Figure 4-3 displays the total data collected for 10 seconds of the baseline and 120 seconds of the task. The focus is on the last 6 seconds of baseline and the first 60 seconds of the task. Two baseline epochs (last 3 seconds and second-to-last 3 seconds), and 11 task epochs (first 3 seconds, 6 seconds sliding windows, with 3 seconds overlap, over the first 30 seconds, and first 60 seconds) are defined and extracted as a data reduction or feature compression strategy. The feature compression strategy results in over 270 raw features per frame, over 20 reduced features per time epoch and over 20 features per task.
Figure 4-3. Thermal signature for baseline and task.

Depending on the specific feature, both raw and epoch-based features are examined for saliency. Therefore temporal sampling features are classified as either raw-temporal features or epoch based features. Spatial sampling involves dividing every frame into 29 non-overlapping segments. Certain segments show more change in the feature over the period of the task as compared to other segments. A graphical user interface (GUI) is developed that allows to visualize the features by selecting a particular subject. For example, in Figure 4-4 after selecting subject number 618, with starting
frame 1 and ending frame 1800, the graph indicates observing the average pixel value for the forehead segment 5. The blue color region of the graph indicates the last 6 seconds (30 fps × 6 seconds = 180 frames) for the baseline pixel intensities, while the green color region indicates the task region for 1800 frames indicating changes in average pixel intensities. The left side of the GUI indicates the whole video for the subject and right side of the GUI indicates the segmentation region for the subject.

Figure 4-4. Feature visualization graphical user interface.
From our NIR data, a method to extract information related to the micro change in the muscle movements for different emotions is developed. A number of facial Action Units (AU) of interest have been defined based on the Ekman Facial Action Coding System (FACS). To date, all of the raw NIR data to generate the 3D data necessary for this task is processed and is currently being assessed and will be reported later.

4.2 3-D Surface features

Image depth information provides an additional dimension to be used for data analysis. The depth information produces more unique features that describe points of interest in an image. The extracted surface features must be simplistic for fast processing, yet representative of the shape of a given surface. The principal curvatures and other shape statistics are used to infer knowledge about the structure of a surface at a given point.

4.2.1 3-D surface data

Unlike other approaches, the Photon-X system uses a single video camera to capture a high-resolution 3D image (as real-time video or individual frames) in various lighting conditions and at any range (determined by lens). To capture images, Photon-X typically uses four megapixel camera sensors. The system captures 3D images by detecting surface normal vectors, whether using the infrared spectrum or the visible light spectrum. It is an extremely high-resolution 3D geometry on a uniform grid using a single camera. The basic data consists of the 3D surface shape, the 3D normal for each pixel on the face and the color image as shown in Figure 4-5.
4.2.2 Principal curvatures

From the differential geometry concepts, the principal curvatures at any given point of a surface are defined by the Eigen values of the shape operator at the point. The principal curvatures measure the surface twist by different amounts in different directions at that point. At a particular point $p$ in 3-dimensional Euclidean space, one may choose a unit normal vector. A normal plane at $p$ is one that contains the normal. It also contains a unique tangent to the surface and cuts the surface in a plane curve. This curve will in general have different curvatures for different normal planes at $p$.

From the normal section, two extreme (maximum and minimum) normal curvatures at each point are defined. These extrema are known as the principal curvatures $k_1, k_2, (k_1 \geq k_2)$, and the corresponding directional vectors, $e_1$ and $e_2$, are the principal directions that are orthogonal. Furthermore, the Gaussian curvature $K$ and the
mean curvature $H$ can then be defined in terms of the principal curvatures and are given as in Equation (4.11)

$$K = k_1 k_2, \quad H = \frac{1}{2}(k_1 + k_2)$$

(4.11)

Figure 4-6(a) shows the 3D surface rendered using OpenGL, Figure 4-6 (b) shows the color-coded depth-map, such that the colors transition from red to blue with distance from the camera. Figure 4-6 (c) shows the 2D image mapped to the 3D surface. Figure 4-6 shows the surface analysis process applied to a 3D face image. Figure 4-6 (d) and Figure 4-6 (e) illustrate the principal curvature values $(k_1, k_2)$ for each pixel of the face image.

Figure 4-6. (a) 3-D Face model obtained (b) depth map of face model (c) 3-D face imagery in frontal view with texture (d) principal curvature $k2$ (e) principal curvature $k1$. 

49
4.2.3 3-D Feature representation using surface primitives

Using these geometrical attributes, every vertex of the 3-D surface is classified into primitive categories. Thus, the geometric surface can be represented in terms of shape primitives which is realized through a mapping process given as in Equation (4.12)

\[ Q: \{K, H, e_1, e_2, k_1, k_2\} \rightarrow \{L_u\}, \quad u = 1, 2, \ldots, U \quad (4.12) \]

Where \( Q \) is a predetermined classification rule and \( \{L_u\} \) is a set of surface categories with \( U \) elements. The result of performing surface primitive classification of the model is shown in Figure 4-7 where \( U \) is set to 8 and \( Q \) is defined by the following rules:

1. Convex (peak) \((k_1 < 0, k_2 < 0)\)
2. Convex cylinder/cone \((k_1 = 0, k_2 < 0)\)
3. Convex saddle \((k_2 < 0 < k_1, |k_1| < |k_2|)\)
4. Minimal surface \((|k_1| = |k_2|)\)
5. Concave saddle \((k_2 < 0 < k_1, |k_1| > |k_2|)\)
6. Concave cylinder/cone \((k_2 = 0, k_1 > 0)\)
7. Concave (pit) \((k_1 > 0, k_2 > 0)\)
8. Planar \((k_1 = 0, k_2 = 0)\)

As shown in Figure 4-7, the face model mostly consists of peak surfaces (as shown in red), concave surfaces (light blue), and minimal surfaces (dark blue). The other shape classes are much less prominent in this case presented in Figure 4-7. From the
classification of surface images, 3-D surfaces can now be formed into a feature vector that is sensitive to structural variances. Such features are effective in discriminating objects with different shape and provide an added layer of information to be used for detection and recognition. Additionally, the surface primitives provide a reduced representation of the image depth information that serves to mitigate noise components and data redundancy. The surface primitives shown in Figure 4-7 give an overview for the feature extraction.

![Surface primitives of depth map.](image)

Figure 4-7. Surface primitives of depth map.

The 3D data features discussed above were collected and analyzed for the threat/challenge response. We do not report the results here as it is still a subject of
evaluation in the efficacy for robust features corresponding to the emotional responses. These 3-D features include convex peak, convex cylinder, convex saddle, minimal surface, concave saddle, concave cylinder, concave pit, planar and these features are represented by 8 different colors as shown in the Figure 4-8.

![Figure 4-8. Surface primitives with 8 different features.](image)

Since the 2D data was determined useful (as detailed in the next section) and explorations of 3D features were investigated, we only report the details for completeness of methods investigated in the research.

4.3 Preliminary analysis using EO features

Various EO features (Forehead energy, Glabella energy, Blink rate, Eye gaze, Saccade and Velocity) are extracted. Several combinations of features are used to determine if they provide any kind of discrimination between threat and challenge subjects. Among all
the features, Glabella energy seems to provide a better classification rate of 68% using a K-Nearest Neighbor classifier with (K = 3 neighbors). The Glabella energy is tracked over a series of sub sampled frames for threatened and challenged samples and is shown in Figure 4-9 and Figure 4-10. The Glabella energy for the threat and challenge participants is varying with no consistent pattern for either of the classes. None of the EO features extracted have been helpful in discriminating threat and challenge participants. So future analysis of threat and challenge data would not include EO features. Note: the lack of confirmed EO results contributes to the literature and the need for thermal imaging solutions. Next we describe a multi-modal sensing strategy for feature analysis for distinguishing threat/challenge responses.
Figure 4-9. Glabella energy tracking for threatened participants.
Figure 4-10. Glabella energy tracking for challenge participants.

4.4 Multimodal feature framework

Figure 4-11 presents an exhaustive list of features that are extracted by modality, and defines the spatial and temporal sampling of each. The result is a large working feature set, from which the best features for human emotional state assessment are selected. Since there are three modalities, a method is proposed to extract features from each category. The 2-D features in grey indicate that those features will not be used in future analysis because none of the features seemed to be able to discriminate the
emotions. The green colored section for the 3-D features indicates that the process is an on-going effort.

![Multi-modal feature framework](image)

**Figure 4-11. Multi-modal feature framework.**

### 4.5 Ground truth estimation from physiological data

Figure 4-12 shows summary data, and the appraisal, emotional, and physiology indexes for the most challenged and most threatened individuals measured during studies 1 and 2, when the person is lying. A comparison of the means and standard deviations reveals that the studies are experienced on average similarly, though there appeared to be positive
emotion and more variability in general for Study 2. Additionally, examining the most challenged persons for these studies, the appraisal ratios are below 1, suggesting that participants believe they can cope with the demanding situation, their positive affect is greater, their negative affect is very minimal, and their physiology is more challenging in Study 2. Unexpectedly, physiology is negative in study 1, perhaps because of the less reliable blood pressure assessment. An examination of the most threatened persons shows higher threat appraisals suggesting that situational demands outweigh coping resources. Further positive affect is lower than the negative affect reports. Lastly, the physiology is less positive, again particularly in study 2.

<table>
<thead>
<tr>
<th>STUDY 1 (n = 20)</th>
<th>STUDY 2 (n = 23)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraisal Ratio</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Positive emotion</td>
<td>1.10 (0.32)</td>
</tr>
<tr>
<td>Negative emotion</td>
<td>2.75 (0.61)</td>
</tr>
<tr>
<td>Physiology</td>
<td>1.62 (0.59)</td>
</tr>
<tr>
<td>Appraisal Ratio</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Positive emotion</td>
<td>1.06 (0.55)</td>
</tr>
<tr>
<td>Negative emotion</td>
<td>3.25 (0.78)</td>
</tr>
<tr>
<td>Physiology</td>
<td>1.66 (0.80)</td>
</tr>
<tr>
<td>Physiologie</td>
<td>0.25 (0.84)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Most Challenged Person:</th>
<th>Most Challenged Person:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraisal Ratio</td>
<td>0.75</td>
</tr>
<tr>
<td>Positive emotion</td>
<td>2.50</td>
</tr>
<tr>
<td>Negative emotion</td>
<td>1.00</td>
</tr>
<tr>
<td>Physiology</td>
<td>-1.02</td>
</tr>
<tr>
<td>Appraisal Ratio</td>
<td>0.24</td>
</tr>
<tr>
<td>Positive emotion</td>
<td>4.40</td>
</tr>
<tr>
<td>Negative emotion</td>
<td>1.00</td>
</tr>
<tr>
<td>Physiology</td>
<td>1.58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Most Threatened Person:</th>
<th>Most Threatened Person:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraisal Ratio</td>
<td>1.82</td>
</tr>
<tr>
<td>Positive emotion</td>
<td>2.30</td>
</tr>
<tr>
<td>Negative emotion</td>
<td>2.90</td>
</tr>
<tr>
<td>Physiology</td>
<td>0.54</td>
</tr>
<tr>
<td>Appraisal Ratio</td>
<td>2.30</td>
</tr>
<tr>
<td>Positive emotion</td>
<td>3.50</td>
</tr>
<tr>
<td>Negative emotion</td>
<td>4.40</td>
</tr>
<tr>
<td>Physiology</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Figure 4-12. Summary data for participants for study 1 and study 2.

Next, an examination of the most challenged and threatened person for each study
shows that for Study 1, the range of appraisals is 0.75 to 1.82, but for Study 2 the range is greater - 0.24 to 2.30. This pattern of greater range for study 2 is similar for positive emotion, negative emotion, and physiology. These data suggest that algorithm development for Study 2 should yield greater separation of threatened and challenged participants.

The physiology ground truth (GT) refers to the combined cardiac output and vascular resistance measures, whereas the appraisal GT is based on self-report (survey) responses. For the appraisal GT, an appraisal ratio is created by first averaging the items that measure primary appraisal, next averaging the items that measure secondary appraisal and, finally, computing the ratio of those two intermediate variables (the ratio of primary average to secondary average) to create the appraisal ratio score. Higher values denote a greater threat and lower values a challenge response.

4.5.1 Data partitioning using physiological GT

The computed physiological Z-scores for Study 1 participants ranged from -2.5 to 2.5 as shown in Figure 4-13. An arbitrary normalized threshold of ± 0.5 to separate “challenged” from “threatened” participants data is selected. Participants for whom the Z-scores fall between -0.5 and 0.5 are labeled “neutral”. Using this partitioning scheme, eight participants fall into the Threatened category (Subjects 514, 504, 541, 519, 531, 543, 548, 539), eight participants fall into the Challenged category (538, 545, 530, 526, 550, 540, 532, 503) and the remaining 25 participants fall into the Neutral category. Study 1 participants were identified as belonging to one of three categories: Threatened
(z-score < -0.5), Neutral (-0.5 < z-score < 0.5) or Challenged (z-score > 0.5).

Figure 4-13. Physiological ground truth.

4.6 Dimensionality reduction for Study 2

The data partitioning for Study 2 based on appraisal ratio ground truth is shown in Figure 4-14. A decision has been made to use seven threat and challenge data samples for classification purposes. The thermal data captured for the “false behavior study” showed some interesting patterns and the feature extraction and classification for the particular study is explored below. The appraisal ratio ground truth is based on self-report (survey) responses and used in our findings. The computed z-scores ranged from -1.5 to 3.0 as shown in Figure 4-14. Higher z-scores indicate “Threat” and lower z-scores indicate
“Challenged”. A minimum of 10 threat and 10 challenge examples are selected in the respective boxes below for classification purposes.

![Figure 4-14. Using appraisal ratio ground truth, participants are divided into two categories: threatened (z-score>0) and challenged (z-score<0).](image)

### 4.6.1 Principal component analysis

Principal component analysis is a statistical method which helps in reducing the dimensionality of the data. PCA is an orthonormal transformation used to transform the number of features into principal components. The principal components are ordered in such a way that the variance is highest for the first principal component and it reduces for the remaining components. Using the first three principal components, in most cases would capture 95% of the variance in the data.

Let \( \mathbf{Y} \in \mathbb{R}^Q \) be represented as a linear combination of orthonormal basis vectors
\( \varphi_1, \varphi_2, \ldots, \varphi_Q \) as in (4.13).

\[
Y = \sum_{i=1}^{Q} X_i \varphi_i \tag{4.13}
\]

Assuming the vector \( Y \) needs to be represented with only \( P \) \((P < Q)\) basis vectors. The way to accomplish the above would require the replacement of the components \([X_{P+1}, \ldots, X_Q]^T\) with some pre-selected constants \( b_i \). The resulting Equation (4.14) would be

\[
Y(P) = \sum_{i=1}^{P} X_i \varphi_i + \sum_{i=P+1}^{Q} b_i \varphi_i \tag{4.14}
\]

The representation error is then denoted as

\[
\Delta Y(P) = Y - Y(P) = \sum_{i=P+1}^{Q} (X_i - b_i) \varphi_i \tag{4.15}
\]

A representation of the error using mean-squared magnitude of \( \Delta Y(P) \) is used. The goal is to minimize the mean-square error (MSE), \( \varepsilon \).

\[
\varepsilon^2(P) = E[|\Delta Y(P)|^2] = E \left[ \sum_{i=P+1}^{Q} \sum_{j=P+1}^{Q} (X_i - b_i)(X_j - b_j) \varphi_i^{T} \varphi_j \right] \tag{4.16}
\]

\[
\varepsilon^2(P) = \sum_{i=P+1}^{Q} E[(X_i - b_i)^2] \tag{4.17}
\]

To find the optimal values of \( b_i \), partial derivative is calculated and equated to zero.
The discarded dimension is replaced by their expected value. The mean square Equation 4.17 can be rewritten after the substitution

\[
\varepsilon^2(P) = \sum_{i=P+1}^Q E[(X_i - E[X_i])^2] \tag{4.19}
\]

and with further re-substituting

\[
\varepsilon^2(P) = \sum_{i=P+1}^Q E[(Y \varphi_i - E[Y \varphi_i])^T (Y \varphi_i - E[Y \varphi_i])] \tag{4.20}
\]

Simplifying the above Equation (4.20) results in

\[
\varepsilon^2(P) = \sum_{i=P+1}^Q \varphi_i^T E[(Y - E[Y])^T (Y - E[Y])] \varphi_i \tag{4.21}
\]

Simplifying the above Equation (4.21) yields,

\[
\varepsilon^2(P) = \sum_{i=P+1}^Q \varphi_i^T \sum_Y \varphi_i \tag{4.22}
\]

where \( \sum_Y \) is the covariance matrix.

The solution that minimizes Equation 4.19, subject to the orthonormality constraint is conducted using a set of Lagrange multipliers \( \lambda_i \).

\[
\varepsilon^2(P) = \sum_{i=P+1}^Q \varphi_i^T \sum_Y \varphi_i + \sum_{i=P+1}^Q \lambda_i (1 - \varphi_i^T \varphi_i) \tag{4.23}
\]

and computing the partial derivative with respect to the basis vectors and equating it to
Equation 4.21 concludes that $\varphi_i$ and $\lambda_i$ is the Eigen vectors and Eigen values of $\Sigma_Y$; respectively. Therefore, to represent $Y$ with minimum MSE, choose the Eigen vectors $\varphi_i$ corresponding to the largest Eigen values $\lambda_i$. The optimal approximation of a random vector $Y \in R^O$ by a linear combination of $P<Q$ independent vectors is obtained by projecting $Y$ onto the Eigen vectors $\varphi_i$ corresponding to the largest Eigen values $\lambda_i$ of the covariance matrix $\Sigma_Y$.

To visualize and quickly interpret the histogram-based feature data, a Principal Component Analysis (PCA) is performed and the data is reduced into its two main dimensions (Principal Components). Figure 4-15 shows that threatened samples (T) and challenged samples (C) form close clusters, and there are only a couple of outliers that spoil perfect clustering. In Figure 4-16, both threatened and challenged samples (T + C) seemed to be separated from Neutral (N) samples. Visualizing data this way helps us to design a simple classifier that may outperform artificial neural networks (ANNs). This approach of examining data is pursued in the further analysis.
Figure 4-15. PCA results of threatened (T) vs. challenged (C) participants.

Figure 4-16. PCA results of threatened + challenged (T+C) vs. neutral (N) participants
4.7 Classification

The classification approach, as diagrammed in Figure 4-17, begins with classifier modules, followed by data sets, different ground truths and, eventually, feature sets. Although additional classifiers will be tested, to date only two strategies are tested: K-Nearest Neighbors and Artificial Neural Networks (ANN). Each of these classifiers has been applied to data from Study 1 and Study 2 using each of the two ground truths (GTs) and inputting visible and thermal features, one feature at a time.

![Figure 4-17. Top-down approach to testing various classification schemes.](image)

4.7.1 Classification using Artificial Neural Network

With the goal of quickly looking at patterns in the dataset and considering the non-linear
behavior of the physiological GT, an artificial neural network (ANN) is implemented. In this case, a two-class problem is presented, using data from the Challenged and Threatened participants and excluding the Neutral participants. The initial classifier was a simple single layer perceptron (SLP) ANN and a moderate classification success is achieved. By allowing additional hidden nodes (~100) in a two-layer feed-forward network (multilayer perceptron or MLP) with a sigmoid function as a threshold in the hidden layer as shown in Figure 4-18. Using the MLP, classification results are improved. In these early tests, the feature vector comprised of the histogram features with 16 windows, six bins per window, and one sampled image per 58 frames. The resultant feature vector leads to 5,568 input nodes that must be reduced to two output nodes (as this is a two-class problem). The ANN was trained using the scaled conjugate gradient back-propagation algorithm.

![Pattern Recognition Neural Network](image)

**Figure 4-18.** A block diagram of the multi-layer perceptron neural network.
Bootstrapping [56] involves partitioning data into separate training and testing datasets. A two-fold cross validation (one training set and one test set) is used. During ANN training, the network is adjusted to minimize classification error in the training set. Once the network is trained, the test data set is classified to provide an independent measure of network performance as denoted by the confusion matrix, as illustrated in Figure 4-19 that lists the classification rate for each class. With only three Threatened and three Challenged participants in the test set, only four discrete success rates (0%, 33%, 67% and 100%) are possible. For the MLP ANN using histogram features, the average classification rate was 66.7%.

![Confusion Matrix](image)

**Figure 4-19.** Confusion matrix for the artificial neural network classifier using histogram feature.
Classification based on ANNs and K-NN did provide some class discrimination, but after iterating through several data partitioning schemes and using different ground truths, it is found that a larger sample size is necessary for better performance. With only seven examples for each class that must be further subdivided into training and test sets, there are not enough examples for effective training of an ANN. A simpler classifier such as K-Nearest Neighbors (K-NN) on a lower dimension feature space will be used for further evaluation.

4.7.2 Classification using K-NN

The K-nearest neighbor is a non-parametric method for classification that predicts objects based on the $k$ closest training examples in feature space. It is one of the simplest of all machine-learning algorithms and is not reported here as it is available in many textbooks. In the next chapter discussing results, a K-NN classification is performed with $k=3$ throughout different experiments. For any of the experiment, first the Eigen analysis is performed and the feature vector is reduced to three dimensions and uses those dimensions for K-NN classification.

4.8 Summary

This chapter presented the feature extraction and classification techniques for representing different emotional states. The ground truth estimation based on physiology and appraisal was also discussed. Eigen analysis using the first three principal components was used for dimensionality reduction. Finally, two different classification techniques with preliminary analysis were presented. These techniques and their
elaborate analysis in the study will be discussed in the next chapter.
CHAPTER V

EXPERIMENTAL RESULTS AND DISCUSSION

The analysis pertaining to this dissertation research involved several different aspects to be considered. There are three different modalities of data collected (Visible, Thermal and 3D datasets). The analysis has been mainly focused on thermal data. Statistical thermal features seemed to extract information that would allow us to discriminate between threat and challenge participants. Even in the case of statistical features, the data is divided into time series (sub-sampling every few frames) or epoch based intervals (averaging over few frames). Since the data set is spatio-temporal, a dimensionality reduction using Eigen analysis is performed which further helps in classification. Artificial neural nets and basic K-Nearest Neighbors are used for discriminating purposes.

5.1 False opinion (study 1) vs. false behavior (study 2) studies
As mentioned earlier in Chapter 3, two deception studies were conducted. A “false opinion study” and a “false behavior study” which both serves as high stake stressors and were collected to understand the facial changes with respect to the human stress states.
For both studies, blood pressure assessment was made on the non-dominant arm. For the first study, an Oscar Ambulatory BP Monitor was used, obtaining two assessments at baseline, one pre-task and two during the task. For study 2, a continuous noninvasive arterial blood pressure monitor (CNAP) was used. Study 2 was similar in many ways to Study 1, in that the measures for physiology, psychological self-report and the multimodal imaging remained the same. However, the nature of the deception differed as shown in Table 5-1.

**Table 5-1. Summary data for lie condition for study 1 and study 2**

<table>
<thead>
<tr>
<th></th>
<th>STUDY 1 (n = 20)</th>
<th>STUDY 2 (n = 23)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Appraisal Ratio</td>
<td>1.10 (0.32)</td>
<td>1.06 (0.55)</td>
</tr>
<tr>
<td>Positive emotion</td>
<td>2.75 (0.61)</td>
<td>3.25 (0.78)</td>
</tr>
<tr>
<td>Negative emotion</td>
<td>1.62 (0.59)</td>
<td>1.66 (0.80)</td>
</tr>
<tr>
<td>Physiology</td>
<td>-0.11 (0.80)</td>
<td>0.25 (0.84)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Most Challenged Person:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraisal Ratio</td>
<td>0.75</td>
<td>0.24</td>
</tr>
<tr>
<td>Positive emotion</td>
<td>2.50</td>
<td>4.40</td>
</tr>
<tr>
<td>Negative emotion</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Physiology</td>
<td>-1.02</td>
<td>1.58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Most Threatened Person:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraisal Ratio</td>
<td>1.82</td>
<td>2.30</td>
</tr>
<tr>
<td>Positive emotion</td>
<td>2.30</td>
<td>3.50</td>
</tr>
<tr>
<td>Negative emotion</td>
<td>2.90</td>
<td>4.40</td>
</tr>
<tr>
<td>Physiology</td>
<td>0.54</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Table 5-1 shows summary data, and the appraisal, emotional, and physiological
indexes for the most challenged and most threatened individuals measured during Studies 1 and 2, for the Lie condition. There appeared to be more variability in self-reports of appraisal and emotional state for Study 2. The most challenged person appeared to have higher positive affect, lower negative affect, and challenge-like physiology than in Study 1. A similar examination of the most threatened persons shows higher threat appraisals, lower positive affect, and higher negative affect, although the physiology appears similar for the threat person in studies 1 and 2 for the threatened person. Physiology metrics appear more disparate and in the right directions for challenged and threatened participants within Study 2. Physiology may have been less reliable for Study 1 given the less precise measurement of blood pressure.

Given the above analysis of appraisal ratio, physiology for studies, study 2 and appraisal ratio as ground truth was the preferred one for reporting the novel analysis in this dissertation. The research and findings in past literature have shown that appraisal ratios distinguish important outcomes for people along the threat-challenge continuum [12]. Preliminary facial pattern analysis for both the studies were performed and concluded that the study 2 showed interesting results and discriminatory analysis using appraisal ratio as ground truth which agreed with the research result and also with Table 5-1.

5.2 Data partitioning using appraisal ratio

As mentioned in Chapter 4, the analysis for Study 2 is performed using appraisal ratio ground truth. The appraisal ratio ground truth is divided into threat, challenge and neutral
participants as shown in the Figure 5-1. This includes all the participants. The computed z-scores ranged from -2.0 to 3.0 as shown in Figure 23. Higher z-scores indicate “threat” and lower z-scores indicate “challenge”. Ten threat and challenge participants as shown in the respective boxes are selected. “Neutral” participants are the ones whose z-scores fall between -1.0 and 1.0. Using this partitioning scheme, ten participants fall into the Threatened category and ten participants fall into the Challenged category and the remaining participants fall into the Neutral category (or less challenged and less threatened category).

Figure 5-1. Appraisal ratio for study 2 participants.
5.3 Statistical feature analysis for threat and challenge participants

From the preliminary analysis, it was observed that the statistical feature representing the mean of the top 10% hottest pixels in a segment is able to distinguish the basic emotional states. The mean of top 10% hottest pixels feature is used in most of the analysis as it was determined to be robust. This feature can be tracked over the whole emotional time cycle starting with baseline and first 30 seconds of the task for all the 29 segments of the face. The time series analysis and the epoch based analysis for this particular feature is described below.

5.3.1 Time series (sub-sampling every few frames)

The first feature selected is the mean of the top 10% hottest pixels and it is tracked over the baseline and the task as shown in Figure 5-2. The baseline for the last 10 seconds consists of 300 frames, while the task is extracted for 60 seconds during a test period of 120 seconds. This provides 1800 frames for analysis. For a particular frame, which contains 29 facial segments and each facial segment contains the mean of top 10% hottest pixels feature. The average of the baseline feature is subtracted from the features of each task frame. The resulting task frame feature vectors are baseline subtracted. Now each frame will consist of 29×1 = 29 features for 29 segments as shown in Figure 5-3. Time series would involve picking a frame of data every second, so it will end up sub-sampling from 900 frames every 30th frame which will result in 30 frames (30 distinct frames × 29 features/frame × 1 feature = 870 feature points /subject). So the feature vector will be
870 features × 20 subjects.

Figure 5-2. Mean of top 10% hot pixel feature tracked over the baseline and task.
5.3.2 Epoch based (averaging over few frames) features

The same feature (mean of top 10% hottest pixel) is again considered for calculating the epochs. The baseline is considered as last ten seconds as shown in Figure 5-4. The baseline epoch that is frequently used for calculations is B031 (which indicates the last 3 seconds of the baseline). The task is divided into windows of 6 seconds interval with 3 seconds sliding window, so the first 30 seconds of the task is divided into 9 epochs starting with T061-----T069. The next step involves subtracting each of the 9 epochs from the baseline B031. For a particular frame of data there are 29 facial segments. For each segment a mean of the top 10% hottest pixel is considered as the dominant feature. This dominant feature will be divided into 9 epochs, so the total features, would be 29 segments/frame × 9 epochs × 1 feature= 261 feature points per participant. The total
5.4 Dimensionality reduction

The time series features represent 870 feature points per participant, while the epoch based features represent 261 features per participant. Certain segments of the face, especially the forehead and nose segments (1, 2, 3, 4, 5, 6, 14, 15 16, 19) are considered, the time series features will be \(30 \text{ frames} \times 10 \text{ segments/frame} = 300 \text{ features/participant}\). Similarly the epoch based features for nose and forehead segments will be \(9 \text{ epochs per frame} \times 10 \text{ segments/frame} = 90 \text{ features/participant}\). The dimensionality reduction of the
features in a 2-D space would help us to visualize the data and would help us to define a classifier. An Eigen analysis is performed for both the type of features. In Eigen analysis, the principal component 1 (PC 1) captures the lowest frequency information, while the principal component 2 (PC 2) and principal component 3 (PC 3) capture higher frequency information. After further analysis, it appeared that the facial changes that happen during the threat or challenge situation are mainly high frequency changes and both the PC2 and PC3 capture them very reliably. The 2-D plot of PC2 and PC3 for time series and epoch based features are as shown in Figure 5-5 and Figure 5-6 respectively.

Figure 5-5. Time series based Eigen analysis.
Figure 5-5 and Figure 5-6 are key contributions of this research as the 2-D plot of PC2 and PC3 for time series and epoch based features show good distinguishing results for threat/challenge responses. The blue participants are the challenged and the red participants are the threatened ones. The comparison of both the plots shows that the separation seems to be similar; in fact same participants are placed in similar position in the plot. The time series based features are reduced from 300 features to 2 features and epoch based is reduced from 90 features to 2 features. Note that there is good separation except for subject 607, which could be an artifact of the data collection and might be removed. Both the time series based and epoch based analysis are observed to be providing similar results. The epoch based features discriminates the dataset with less
features as compared to time series and therefore, in all the future analysis, epoch based features are used.

To understand why PC 2 and PC 3 provide a better discrimination, a bar chart representation is considered. A plot of principal component versus the variance is shown in Figure 5-7. The first three principal components seemed to capture 95% of the variance in the data. However, even though the PC 1 captures 65% of the total variance, it does not seem to help with the feature discrimination.

Figure 5-7. Principal component vs. variance.
To understand the contribution of each feature, a biplot is used for visualization purposes. A biplot allows visualizing the magnitude and sign of each variable’s contribution to the first two or three principal components and how each observation is represented in terms of those components. Figure 5-8 presents a biplot for the PC 1 and PC 2. The observations are distributed all over the biplot, however, each feature has contribution on the positive side of the PC 1 and no contributions on the negative side of the PC 1. In contrast, Figure 5-9 presents a biplot for the PC 2 and PC 3 where the observations are distributed all over the biplot, and each feature has contribution in all the four quadrants of the biplot. Figure 5-8 and Figure 5-9 show that the features contribution presented by PC 2 and PC 3 are distributed uniformly and classify the data set robustly.

Figure 5-8. Biplot describing PC 1 vs. PC 2 for epoch based features.
Next, an investigation is performed to analyze the principal component to better understand the data and classification method. The principal component loadings are correlation coefficients between the principal component scores and the original variables. The loadings can also be defined as the weight by which each standardized original variable should be multiplied to get the component score. PC 1 determines level or tilt of the dataset, PC 2 determines slope or trend of a particular dataset, PC 3 signifies curvature or twist, as per the research paper [57]. Visually it seems like there is a great change of slope in the dataset as shown in Figure 5-10, where PC 2 seems to capture and in turn may help to classify both classes. When the plot of the loadings for all three PCs are plotted, the PC 1 had all positive loadings indicating level, PC 2 did change sign for
the loadings once indicating change in slope and PC 3 did change sign twice indicating the curvature. As mentioned in the section 5.4, a total of 90 epoch based features would result in 90 principal components after the Eigen analysis. A feature selection technique with objective function defined as classification rate is used on all the principal components and found that PC 2 and PC 3 give the best classification rate. This reconfirms that PC 2 and PC 3 have better discrimination capability than PC 1. PCA does not consider the class separability since it does not take into account the class label. There is no guarantee that direction of maximum variance will contain good features for discrimination. Analyzing the raw data, it appears that for threatened examples there is a positive change of slope from feature 9 to feature 10, while for challenged examples there is a negative change of slope from feature 9 to feature 10. Additional analysis of the raw data is considered in the Section 5.9.
5.5 Classification for thermal data

A K-Nearest Neighbor scheme with \( k = 3 \) neighbors is used for classification. The face region is divided into six sub-regions as described earlier. Forehead and nose region together provide a very distinct classification as shown in Figure 5-11. The K-NN classification rate for the nose segments is 80%, while the K-NN classification rate for forehead separately is 70%. Fusing both the forehead and nose segments would increase the classification rate. The K-Nearest Neighbor classification rate of 85% is achieved discriminating threat and challenge examples in the forehead and nose regions. Feature selection illustrated in Figure 5-11 shows that the forehead and nose regions of the face
are more useful than other regions of the face, which is consistent with other research findings [43].

Figure 5-11. Scatter plots with PC 2 and PC 3 for forehead, nose and fusion of forehead and nose, classification accuracy using K-NN for forehead, nose and fusion are 70%, 80% and 85% respectively.

5.6 Region based classification
The face region is divided into six distinct regions with specific segments denoted with explicit colors as shown in Figure 5-12. An epoch based statistical features were used to further reduce to specific transformed features using Eigen based analysis. To visualize and quickly interpret the epoch based feature data, an Eigen based analysis is performed.
and reduced the data into its two main dimensions and then performed K-NN classification on each segment of the data. Certain regions of the face together provide a better discrimination than others. An elaborate combination of all regions and feature selection of particular region is performed. All 6 regions encompassing 29 segments (whole face) did undergo an Eigen analysis and it was followed by K-NN classification. The classification rate did drop to 73% as mentioned in Table 5 for the entire face. The classification rate pertaining to nose and forehead feature selection shows higher classification rates as shown in Table 5-2.

![Face segmentation showing forehead (green), eye-periorbital (cyan), cheek (yellow), nose (red), mouth (blue), and neck (magenta).](image)

Figure 5-12. Face segmentation showing forehead (green), eye-periorbital (cyan), cheek (yellow), nose (red), mouth (blue), and neck (magenta).
Table 5-2. Region based classification.

<table>
<thead>
<tr>
<th>Region</th>
<th>Segments</th>
<th>Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forehead</td>
<td>1, 2, 3, 4, 5, 6</td>
<td>80%</td>
</tr>
<tr>
<td>Nose</td>
<td>14, 15, 16, 19</td>
<td>70%</td>
</tr>
<tr>
<td>Eye</td>
<td>7, 8, 9, 10, 11</td>
<td>50%</td>
</tr>
<tr>
<td>Cheek</td>
<td>12, 13, 20, 23</td>
<td>55%</td>
</tr>
<tr>
<td>Mouth</td>
<td>21, 24, 25</td>
<td>35%</td>
</tr>
<tr>
<td>Neck</td>
<td>27, 28, 29</td>
<td>30%</td>
</tr>
<tr>
<td>Face (Region fusion)</td>
<td>[1 --- 29]</td>
<td>65%</td>
</tr>
<tr>
<td>Forehead + Nose</td>
<td>1, 2, 3, 4, 5, 6, 14, 15, 16, 19</td>
<td>85%</td>
</tr>
</tbody>
</table>

5.7 Pattern analysis for different regions of the face with varying epochs

A series of experimental analysis was performed for different regions of the face including the whole face for varying epochs. This method is an experimental way to select different face sub-regions for improving the classification of the overall system. This method also helps in determining the duration of specific heat patterns that does uphold once the emotion is activated. A set of eight different experiments is performed to understand the importance of specific epochs and specific regions of the face for classification purposes.

In each of the experiment, a total of 20 participants are used, of which 10 participants
belonged to threat category and remaining 10 belonged to challenge category. The feature used for the experiment is the mean of top 10% hot pixel statistical feature for each facial region. A leave one out cross validation strategy is used for performing kNN classification. Finally, accuracy is calculated for three different cases: In the first case accuracy is estimated using all the principal components (PCs), while in the second case first 3 PCs are used for estimating accuracy. Finally, the third case involves using the second and third PC to estimate the accuracy. For each case a confusion matrix is calculated where the ‘T’ indicates threat value, ‘C’ indicates challenge value and ‘N’ indicates that the predicted value does not belong to ‘T’ or ‘C’ and therefore, it belongs to neutral category or less threatened and less challenged category. The rows of the confusion matrix indicate predicted class while the columns of the confusion matrix indicate actual class.

5.7.1 Pattern analysis experiment I: whole face region + all the epochs

The statistical feature (mean of top 10% hot pixels) is estimated for all the 29 segments of the face and for all the 9 epochs (30 seconds) of the particular emotion. The feature vector would be 29 sub-regions x 9 epochs x 1 feature=261 values row vector for each participant. The accuracy is calculated for three different cases as shown in Figure 5-13. The accuracy calculated for each case indicates that PC 2 and PC 3 provide a better separation for the threatened and challenged data indicating that the thermal changes on the face are mostly higher order frequency changes.
5.7.2 Pattern analysis experiment II: whole face region + first two epochs

The statistical feature (mean of top 10% hot pixels) is estimated for all the 29 segments of the face and for the first two epochs (9 seconds) of the particular emotion. The feature vector would be 29 sub-regions x 2 epochs x 1 feature=58 values row vector for each participant. The accuracy is calculated for three different cases as shown in Figure 5-14. The accuracy calculated for each case indicates that PC 2 and PC 3 provide a better separation for the threatened and challenged data indicating that the thermal changes on the face are mostly higher order frequency changes.
Figure 5-14. Accuracy estimation along with confusion matrices for the whole face with first two epochs.

5.7.3 Pattern analysis experiment III: Forehead region + all epochs

This statistical feature (mean of top 10% hot pixels) is estimated for all the 6 segments of the forehead and for the all the 9 epochs (30 seconds) of the particular emotion. The feature vector would be 6 sub-regions x 9 epochs x 1 feature = 54 values row feature vector for each participant. The accuracy is calculated for three different cases as shown in Figure 5-15. The accuracy calculated for each case indicates that PC 2 and PC 3 provide a better separation for the threatened and challenged data indicating that the thermal changes on the face are mostly higher order frequency changes.
5.7.4 Pattern analysis experiment IV: Forehead region + first two epochs

The statistical feature (mean of top 10% hot pixels) is estimated for the 6 segments of the forehead and for the first two epochs (9 seconds) of the particular emotion. The feature vector would be 6 sub-regions x 2 epochs x 1 feature = 12 values row vector for each participant. The accuracy is calculated for three different cases as shown in Figure 5-16. The accuracy calculated for each case indicates that PC 2 and PC 3 provide a better separation for the threatened and challenged data indicating that the thermal changes on the face are mostly higher order frequency changes.
Figure 5-16. Accuracy estimation along with confusion matrices for the whole face with first two epochs.

5.7.5 Pattern analysis experiment V: Nose region + all epochs

The statistical feature (mean of top 10% hot pixels) is estimated for the 2 segments of the nose and for all the 9 epochs (30 seconds) of the particular emotion. The feature vector would be 2 sub-regions x 9 epochs x 1 feature=18 values row feature vector for each participant. The accuracy is calculated for three different cases as shown in Figure 5-17. The accuracy calculated for each case indicates that using two principal components (PC 2 and PC 3) could provide an accuracy of 75% and adding more principal component does not improve the accuracy.
5.7.6 Pattern analysis experiment VI: Nose region + first two epochs

The statistical feature (mean of top 10% hot pixels) is estimated for all the 2 segments of the nose and for the first 2 epochs (9 seconds) of the particular emotion. The feature vector would be 2 sub-regions x 2 epochs x 1 feature=4 value row feature vector for each participant. The accuracy is calculated for three different cases as shown in Figure 5-18. In this case, the accuracy calculated for PC 1, 2, 3 together exceeds the accuracy calculated using PC 2 and PC 3, which indicates that PC 1 does play an important role in discrimination and the lower frequency information in the nose region is also helpful for discrimination. Also the original feature vector is four features and applying Eigen analysis to reduce that small a feature size to even smaller feature size makes no practical sense.

---

Figure 5-17. Accuracy estimation along with confusion matrices for the nose with all the epochs.
Figure 5-18. Accuracy estimation along with confusion matrices for the nose with first two epochs

5.7.7 Pattern analysis experiment VII: (Forehead + Nose) region + all epochs

The statistical feature (mean of top 10% hot pixels) is estimated for all the 8 segments of the forehead and nose and for all the epochs (30 seconds) of the particular emotion. The feature vector would be 8 sub-regions x 9 epochs x 1 feature=72 values row feature vector for each participant. The accuracy is calculated for three different cases as shown in Figure 5-19. The accuracy calculated for each case indicates that PC 2 and PC 3 provide a better separation for the threatened and challenged data indicating that the thermal changes on the face are mostly higher order frequency changes.
5.7.8 Pattern analysis experiment VIII: (Forehead + Nose) region + first two epochs

The statistical feature (mean of top 10% hot pixels) is estimated for all the 8 segments of the forehead and nose and for the first two epochs (9 seconds) of the particular emotion. The feature vector would be 8 sub-regions x 2 epochs x 1 feature=16 values row feature vector for each participant. The accuracy is calculated for three different cases as shown in Figure 5-20. The accuracy calculated for each case indicates that PC 2 and PC 3 provide a better separation for the threatened and challenged data indicating that the thermal changes on the face are mostly higher order frequency changes.
Figure 5-20. Accuracy estimation along with confusion matrices for the fusion of nose and forehead for first two epochs.

5.7.9 Experiments summary

Among all the eight experiments, the region with forehead and nose provides the best accuracy for first two epochs as shown in the Table 5-3. This analysis proves that the initial epochs contain the peak of the emotion and is extremely helpful during discrimination. The Eigen analysis denotes that the fused region of the forehead and nose contain high frequency temperature changes and provide relevant information for discrimination.
Table 5-3. Accuracy estimation for all experiments

<table>
<thead>
<tr>
<th>Experiments</th>
<th>Accuracy all PCs</th>
<th>Accuracy PC 1, 2, 3</th>
<th>Accuracy PC 2, 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face (all epochs)</td>
<td>50%</td>
<td>60%</td>
<td>65%</td>
</tr>
<tr>
<td>Face (First 2 epochs)</td>
<td>55%</td>
<td>60%</td>
<td>65%</td>
</tr>
<tr>
<td>Forehead (all epochs)</td>
<td>50%</td>
<td>50%</td>
<td>65%</td>
</tr>
<tr>
<td>Forehead (first 2 epochs)</td>
<td>45%</td>
<td>35%</td>
<td>80%</td>
</tr>
<tr>
<td>Nose (all epochs)</td>
<td>75%</td>
<td>75%</td>
<td>75%</td>
</tr>
<tr>
<td>Nose (first 2 epochs)</td>
<td>75%</td>
<td>75%</td>
<td>70%</td>
</tr>
<tr>
<td>Forehead + nose (all epochs)</td>
<td>60%</td>
<td>65%</td>
<td>70%</td>
</tr>
<tr>
<td>Forehead + nose (first 2 epochs)</td>
<td>60%</td>
<td>60%</td>
<td>85%</td>
</tr>
</tbody>
</table>

5.8 Pattern analysis for test data set with 30 participants for Study 2

The majority of our pattern analysis discussed involved using the extreme threatened and challenge participants. In this analysis, some new participants are included that belong to the category of lesser threatened and lesser challenged class (neutral) as shown in the Figure 5-21. The dotted box in the challenge category includes the original extreme challenge participants and also five new participants that belonged to the lesser challenged class. Similarly, the dotted box in the threat category includes the original extreme threat participants and also five new participants that belonged to the lesser
A total of 30 participants are used for the analysis. In the threat class, 10 participants belonged to extreme threat class and remaining five belonged to lesser threat class. In the challenge class, 10 participants belonged to extreme challenge class and remaining five belonged to lesser challenge class. From the Table 5-3, it is evident that the fused forehead and nose region provides better accuracy with two epochs. Henceforth, fused forehead and nose regions with two epochs are used for most of the analysis. The statistical feature (mean of top 10% hot pixels) is estimated for all the 8 segments of the forehead and nose and for the first two epochs (9 seconds) of the particular emotion. The feature vector would be 8 sub-regions x 2 epochs x 1 feature=16

Figure 5-21. Data partitioning indicating additional (five lesser threat + five lesser challenge) test participants shown in the dotted rectangular box
values row feature vector for each participant. The cross validation involves a previously
trained system with 20 extreme threat and challenge participants and testing involves
using all the 30 participants. Finally, accuracy is calculated for three different cases using
kNN classification as shown in Figure 5-22: In the first case accuracy is estimated using
all the principal components (PCs), while in the second case first 3 PCs are used for
estimating accuracy. Finally, the third case involves using the second and third PC to
estimate the accuracy. The accuracy calculated for each case indicates that PC 2 and PC 3
provide a better separation for the lesser threatened and challenged data indicating that
the thermal changes on the face are mostly higher order frequency changes. However, the
accuracy dropped from 85% to 70% indicating that the lesser threat and challenge
participants probably do not exhibit a characteristic threat or challenge response.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy All PCs</th>
<th>Accuracy PC 1, 2, 3</th>
<th>Accuracy PC 2, 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>53.33%</td>
<td>56.67%</td>
<td>70%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T</th>
<th>C</th>
<th>T</th>
<th>C</th>
<th>T</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.67</td>
<td>0.53</td>
<td>0.6</td>
<td>0.33</td>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>0.27</td>
<td>0.4</td>
<td>0.27</td>
<td>0.53</td>
<td>0</td>
<td>0.6</td>
</tr>
<tr>
<td>0.07</td>
<td>0.07</td>
<td>0.13</td>
<td>0.14</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Figure 5-22. Accuracy estimation along with confusion matrices for the fusion of
nose and forehead for first two epochs for lesser threat and lesser challenge
participants.

99
5.9 Pattern analysis for test data set with 40 participants for Study 2

The majority of our pattern analysis discussed involved using the extreme threatened and challenge participants. In this analysis, additional new participants are included that belong to the category of lesser threatened and lesser challenged class (neutral) as shown in the Figure 5-23. The dotted box in the challenge category includes the original extreme challenge participants and also ten new participants that belonged to the lesser challenged class. Similarly, the dotted box in the threat category includes the original extreme threat participants and also ten new participants that belonged to the lesser threatened class.

![Figure 5-23. Data partitioning indicating additional (ten lesser threat + ten lesser challenge) test participants shown in the dotted rectangular box](image)

A total of 40 participants are used for the analysis. In the threat class, 10
participants belonged to extreme threat class and remaining ten belonged to lesser threat class. In the challenge class, ten participants belonged to extreme challenge class and remaining ten belonged to lesser challenge class. From the Table 5-3, it is evident that the fused forehead and nose region provides better accuracy with two epochs. Therefore, fused forehead and nose regions with two epochs are used for most of the analysis. The statistical feature (mean of top 10% hot pixels) is estimated for all the 8 segments of the forehead and nose and for the first two epochs (9 seconds) of the particular emotion. The feature vector would be 8 sub-regions x 2 epochs x 1 feature=16 values row feature vector for each participant. The cross validation involves a previously trained system with 20 extreme threat and challenge participants and testing involves using all the 40 participants. Finally, accuracy is calculated for three different cases using kNN classification as shown in Figure 5-24: In the first case accuracy is estimated using all the principal components (PCs), while in the second case first 3 PCs are used for estimating accuracy. Finally, the third case involves using the second and third PC to estimate the accuracy. The accuracy calculated for each case indicates that PC 1, PC 2 and PC 3 together provide similar classification as PC 2 and PC 3. However, the accuracy dropped from 85% to 52.50% indicating that the lesser threat and challenge participants probably do not exhibit a characteristic threat or challenge response.
From the subsection 5.8 and 5.9, the accuracy rate drops considerably from 85% to 70% and further drops to 52.5% as the lesser threatened and lesser challenged participants are included in the testing dataset. The pattern analysis agrees with the psychological ground truth where the participants close to zero z-score are the typical neutral participants as shown in Figure 5-1.

5.10 Deception based classification

The results discussed till now were solely based on threat and challenge participants based on appraisal ratio described in Figure 5-1. There are a total of 23 participants as described in Table 5-1 for the deception condition for the Study 2. Among these participants, 5 participants belong to the threatened group and 5 participants belong to the
challenged group and the rest of the participants were categorized in the neutral group as shown in Figure 5-25.

Figure 5-25. Deception based participants in threat, challenge and neutral category.

Therefore, among those randomly assigned to the deception condition in Study 2, they comprise a subset of the threatened and challenged participants for the larger group. As such, similar Eigen analysis for this small deception group revealed good separation between the threat and challenge participants as shown in the Figure 5-26. This being such a small set, the classification rate of 90% using K-NN may not be statistically
significant. The classification findings suggesting that both the deception and the mixed group (deception and non-deception) participants can be easily classified into threatened and challenged groups based on the extracted thermal features.

Figure 5-26. Deception participants scatter plots.

5.11 Sample size versus total number of features

One of the problems was using a large number of features for discriminating very few threat and challenge samples. The sample size (n) is 20 with 10 examples of threat and 10
examples of challenge. A large feature size would eventually over train the data and the classification rate won’t be accurate. So a decision was made to find the optimum number of features to get a good classification rate. One of the methods is going through the feature space and undergoing a feature selection process. However, it was a known fact from the region based classification in Table 5-2 that nose and forehead region together provides the best classification rate. The forehead and nose segments are selected and a sequential forward search (SFS) algorithm. Among the segments initially selected (1, 2, 3, 4, 5, 6, 14, 15, 16, 19), segment 14 and segment 16 near the Glabella region are selected using SFS. The motivation for this analysis was to try to reduce the feature size to square root of sample size without immensely hurting the classification. The total sample size is 20 samples (10 of each threat and challenge). Square root of 20 is the total number of features that are needed to limit the number of features. Square root of 20 is 4.47 which is approximately 4 features that we sought for effective classification.

For the above experiment the two initial epochs T061 and T062 will cover 9 seconds of a task. These two epochs are baseline subtracted. One of the most prominent features is the mean of the top 10% hot pixels. The segments selected for classification are segment 14 and segment 16. The nomenclature used to describe the feature is as follows. The segment number is used from where the features are extracted, followed by the statistical feature itself and then followed by specific epoch. The four features used for this analysis are as follows:

1. Segment14_top_ten_mean_T061
2. Segment14_top_ten_mean_T062
3. Segment16_top_ten_mean_T061
4. Segment16_top_ten_mean_T062

Figure 5-27. Threatened feature data with positive slope.
Figure 5-28. Challenged feature data with negative slope.

A definite increase in slope for all the threatened participants and decrease in slope for challenged participants in Figure 5-27 and Figure 5-28. This slope change can vary depending on alignment of features. However, the patterns for threat and challenge seemed to be in opposite direction. Using K-NN classifier a classification rate of 83.3% was achieved. Thus, we were able to determine a minimum of four features would be needed to classify threatened and challenged emotional states effectively.
5.12 Summary

All the experimental results were evaluated using study 2 and applying appraisal ratio as the ground truth. The time series and the epoch based features provided identical classification results. The feature selection helped in discriminating the threat and challenge data more accurately. Region based classification allowed us to evaluate each region separately. The feature fusion did improve the classification performance. Threat and challenge emotions were characterized by the PC 2 and PC 3 which indicated the changes in the emotions are high frequency variations. We also found the four salient features that distinguished threat and challenge responses.
CHAPTER VI

CONCLUSIONS AND FUTURE WORK

The development of a feature extraction method for distinguishing threat and challenge emotional states from thermal data was presented in this dissertation. Several analyses were performed to determine if specific regions of the face radiate enough temperature change information during a particular human state that could eventually be used to discriminate specific emotions. An epoch based windowing approach was used to track and extract features from the facial thermo-grams, and then an Eigen analysis was applied to help distinguish different emotions. The proposed approach was tested with ten threatened and challenged individuals whose spontaneous changes in heat patterns were captured and validated with psychological self-reporting ground truth information.

One of the objectives in this study was to define areas of the face which would help to discriminate human emotions. Each human state induces certain change of temperature in specific areas of the face. As noticed from the results presented in Chapter 5, the forehead region (frontalis) and the nose region (nasalis) seem to provide discriminatory information for classification. In examining the patterns for threat and challenge, it was
observed that the temperature of the nasalis region decreases for threatened individuals and increases (or has a forward slope) for challenged individuals. The other regions of the face show changes in temperature; however, these changes may be similar for both of the human states and are not helpful in classification. A comparison of feature fusion of different regions was also performed. The performance of the classifier independently using the features from the forehead region and the nose region are separately evaluated. It was observed that the forehead region provided a classification accuracy of 70% and the nose region provided an accuracy of 80%. The fusion of forehead and the nose region features together provided an improved performance of 85% classification accuracy. However, other regions of the face such as cheek, mouth, eye and neck do not seem to have enough discriminatory information to classify threat and challenge emotional states. When all the features of the face are fused, the classification rate decreased by 10%. This proves that only certain regions of the face collected from thermal imagery provide relevant information to distinguish the threatened and challenged participants. In summary, thermal imaging seems to capture temperature changes in the face, which helps to classify distinct human emotional states from stand-off sensors.

This study will be helpful to understand the peak response of an emotion. This can also be used to determine the duration of a specific heat pattern that could uphold once the emotion is activated. The epoch based windowing allows to use specific windows to determine the duration of the emotion or the activation of a specific heat pattern. A classification rate for each window could be computed which will help to decide the peak
or apex of the emotion. In this study, the threat and challenge participants were discriminated using only the thermal inputs. The 3-D NIR camera will be able to provide more complementary information to the thermal camera. Using the 3-D camera it will be possible to relate the action units to different emotions. Research work is progressing to study the effect of a fused representation of 3-D and thermal features for detecting various emotional states.

Future aspects of the research could be used in designing test-beds for human performance and training in surveillance tasks [58]. Also, the classification method from the thermal imaging could be combined with other sensors for classifier fusion [59] or confusion matrix fusion [60].
References


[40] I. Pavlidis, N. L. Eberhardt and J. A. Levine, "Human behaviour: Seeing through the


[52] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere and W. Wu, "Objective assessment


