Detecting internal emotions using video cameras
（ビデオカメラを用いた内面的感情の認識）

KEYA DAS TILOTTOMA

Graduate School of Science and Engineering

Saitama University, Japan

A thesis submitted for the degree of

Doctor of Philosophy

March, 2020

This thesis is dedicated to
Das Yuri, Das Wataru,
Das Anjan
for their endless love and support

Acknowledgements

First, I thank my **PARENTS** for giving me the strength and ability to complete this study. There are also many people who have supported and encouraged me throughout this doctoral research that I would like to acknowledge.

I would like to thank my supervisor, Professor **Yoshinori Kuno**, Saitama University, Japan, for his trust, enormous support, perfect guidance, strong encouragement, patience and for being always there since the beginning. I am sincerely grateful to him for his constant encouragement, and his stimulating ideas.

I am incredibly grateful to Professor **Yoshinori Kobayashi**, and Assistant Professors **Antony Lam** and **Hisato Fukuda** for their comments, insightful discussions, valuable advice, and feedback during evaluating of my thesis work. I am forever thankful to all lab members for their sharing knowledge and their time and help that they have given to me. For this dissertation I would like to thank my oral defense committee members: Professor **Testsuya Shimamura** and Professor **Takashi Komuro** for their time, interest, insightful questions and helpful comments.

Finally, but very certainly not least, I am grateful to my kids Das Yuri and Das Wataru, and my husband, Das Anjan who had to bear with my moods and impatience whenever the research did not work out the way I wanted. Without their love, care, support, and presence, I would not have been able to complete my doctoral study in Japan.

Abstract

Automatically recognizing of human emotion is an interesting and challenging task with many applications such as human robot interaction, movie marketing, and more. Emotion analysis and recognition has become an interesting topic of research among the computer vision research community, Human-Computer Interaction (HCI) and Human-Robot Interaction (HRI) is one of the interesting challenges in the community of human-computer interaction today to make computers be more human-like for intelligent user interfaces. Emotion, one of the users affects, has been recognized as one of the most important ways of people to communicate with each other. Given the importance and potential of the emotions, effective interfaces using the emotion of the human user are gradually more desirable in intelligent user interfaces such as Human-Robot Interactions. Thus, there has been much work on systems that identify emotional states. Human emotion recognition using facial expressions is a common approach and there are many researchers working in this direction. However, there are times when facial expressions can either be faked or hidden. That is, one's "apparent emotions" from say, facial expressions may not be a reflection of one's genuine "inner emotions". Thus, other modalities such as physiological responses should be investigated for detecting and recognizing a person's inner emotions. This thesis aims to build a practical system for detecting such emotions by observing physiological responses. We focus on sensing physiological changes through visual means using only conventional cameras, as this would not require specialized equipment.

As mentioned earlier, the computer vision community has made many advancements in apparent emotion recognition through facial expressions. On the other hand, the psychophysiology community has conducted several studies on detecting and recognizing internal emotions using different physiological channels. Recognition of emotion has been done using many physiological signs such as heart rate change, eye movements, eye blinks,

change in skin conductance, and change of skin temperature. Although many different physiological signs are used, many researchers find that cardiac activity is useful for emotion recognition. As a result, this thesis explores the use of cardiac activity for emotion recognition. However, most past studies use electrocardiography (ECG) for reading cardiac activity. ECGs are effective but have their limitations due to the need for attached sensors and higher cost. To realize practical application system for many real-world settings, we need a method that can sense cardiac activity without requiring any wearable attachments or devices.

Fortunately, in recent years, remote photo plethysmography (PPG) algorithms have received attention. These techniques allow one to read cardiac activity such as heart rates (HR) from conventional cameras by typically observing small changes in skin color over time. This thesis aims to use the sensed cardiac activity from remote PPG to detect and recognize emotions. Since remote PPG has been shown to work with conventional cameras, our proposed approach has the benefit that cameras such as webcams, surveillance cameras, and cellphone cameras could be used. With the ability to see cardiac activity without contact sensors, we present a convenient system for detecting internal emotions. Like in the psychophysiology literature, we chose to evaluate our approach by recognizing emotional reactions to emotionally stimulating videos such as horror and comedy clips. In the first phase of our work, we showed video content to human subjects and collected HR data using an attached sensor (Fitbit) for three emotional states (normal resting, funny, and horror). We then confirmed that the average HR between normal resting states and the emotionally stimulated states exhibit a statistically significant difference.

The first phase of our work showed that HRs could be used to detect changes in emotional state but did not explore the recognition of what kinds of emotions were present. In the next phase of our work, we investigated the feasibility of using cardiac pulse signals for recognizing different emotional states (joy vs. fear) and how this compares with the use of facial expressions. Specifically, we used the Open Face facial landmark tracker to estimate the average facial action unit intensities for each subject on the 30 second segments in both the comedy and horror cases.In this Thesis, we use a remote video based cardiac activity sensing technique

to obtain physiological data to identify emotional states. We show that from the remotely sensed cardiac pulse patterns alone, emotional states can be differentiated. Specifically, we conducted an experimental study on recognizing the emotions of people watching video clips. We recorded all volunteers that all watched the same comedy and horror video clips and then we estimated their cardiac pulse signals from the video footage. From the cardiac pulse signal alone, we were able to classify whether the subjects were watching the comedy or horror video clip. We also compare against classifying for the same task using facial action units and discuss how the two modalities compare. In experimental period all subjects are watching two different kinds of emotional status changes video clips, like comedy and horror video clips. By using their pulse signal, we have analysis and tries to find their emotional status or emotional classification during watching the video clips. We have compared HR method with various types of wearable sensors and features, like Wii fit balanced board, pulse oximeter and GSR. Those wearable sensors are used to detect physiological signal.

In short, the two main contributions of this PhD thesis include sensing emotional changes in humans and determining the effectiveness of emotion recognition using only remote PPG sensed data relative to conventional facial expression analysis. To our knowledge, we are one of the first to bridge the gap between computer vision and psychophysiology through presentation of a promising system for visual detection of internal emotions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Internal Emotions

Human emotions are very important for human life, because it affect physiological and psychological status of people. However, to detect human emotions status is little bit challenging. A popular approach of identifying the emotion is from facial expressions, because facial expressions convey very important symptom of emotions. In general, people infer the emotional states of other people, such as joy, sadness, and anger, using facial expressions. Many researchers work on this field for example, Cohen et al. [1] proposed a method for facial expression recognition from video. They introduced a Tree-Augmented-Naive Bayes (TAN) classifier that learns the dependencies between facial features and they provide an algorithm for finding the best TAN structure. In Zhang et al. [2] they propose a method of facial expression recognition based on local binary patterns (LBP) and local Fisher discriminant analysis (LFDA). However, the main limitation with facial expression recognition systems is that sometimes human face is not show real emotions as shown in Figure 1.1.

In this work, we make the distinction between external and internal emotions. External emotions are readily visible through facial expressions and there is a wealth of work on this topic. But sometimes our emotions cannot show our face. This type of emotions we called internal emotions.

Figure 1.1: Sometimes human show fake emotion

Source: https://www.telegraph.co.uk/technology/apple/12088601/Apple-buys-company-that-scans-your-face-to-read-emotions.html

1.2 Internal Emotion Recognition from Heart Rate

Sometimes our face may not show our internal or real emotions. That's why we need to know our internal or real emotions. When we know human internal emotions or real emotions that means we know their real demands and acutely they want. Many researchers work on emotion recognition from human physiological signal, that has been a hot topic recent year. John Stern defined psychophysiology as "any research in which the dependent variable (the subject's response) is a physiological measure and the independent variable (the factor manipulated by the experimenter) a behavioral one" [3].Moreover, it has also been reported that subjects have experienced emotions, which were manifested through physiological responses without showing any visible changes in facial expressions [18]. Indeed, emotions can be defined as a mental state that occurs spontaneously without any conscious effort and accompanied by physiological changes [18]. Many kinds of physiological signals have been applied to emotion recognition, including electrocardiogram (ECG), galvanic skin response (GSR), electroencephalogram (EEG), respiratory suspended particulate (RSP) blood volume pulse (BVP) heart rate (HR), skin conductance level (SCL) and skin conductance response (SCR). These physiological signals, specially heart rate is very popular for recognition emotions.

Many Researchers work for emotion recognition based on heart rate but they have been using an electroencephalogram (ECG). There are many examples of work in this area. For example, Kim et al. [3] reported an emotion recognition system with 78.4% and 6l.8% accuracy for the recognition of 3 and 4 classes of emotions using ECG, skin temperature variation, and electrodermal activity. Zong et al. [4] used 25 features from ECG, electromyogram, skin conductivity (SC), and respiration changes by the Hilbert-Huang transform to obtain 76% accuracy for 4 classes. Picard et al. [5] used 40 features from heart rate, muscle tension, temperature, and SC to get 81% recognition accuracy on 8 classes. Guillaume et al. [6] obtained 80% recognition accuracy on 3 classes using an electroencephalogram (EEG). These researchers reinforce the finding that physiological changes primarily respond to emotion.

1.3 Heart Rate from Video

The ability to use conventional camera store motley measure heart rate (HR) would open doors to many possibilities. For example, HR measurement using cameras could provide emotion recognition system, which would be of great benefit to effective computing in human-computer and human-robot interaction applications. Another potential application would be in health monitoring. A networked camera system could be set up in a nursing home to continually monitor patient health for the long term without uncomfortable sensors. Real time monitoring of HR over web cam could also be done by medical sectors. Video can represent emotions and be an emotion inductor too.

In this paper we consider heart rate (HR) and its connection to emotional states. However, conventional methods for measuring HR such as ECG or photoplethysmography (PPG) using optical sensors require the sensor to make physical contact with the person and this can be inconvenient and uncomfortable for the person. Fortunately, in controlled conditions it has been shown that it is possible to use a conventional camera to remotely detect small changes in skin color due to a person's cardiac pulse [13]. Verkruysse et al. [11], found that conventional RGB cameras such as the Canon Power shot can capture small changes in light absorption that correspond well with the cardiac pulse. However, their tests were conducted under controlled conditions. Later, several authors extended their work to account for more realistic settings [7,10,12,13,15]. Thus, we use a video PPG method. Therefore, our goal is to use the cardiac pulse signal (PPG signal) to perform emotion recognition. So, what we do it that after estimating the final HR, we go back and identify all the PPG signals that contributed to the final HR estimate. We then normalized each of those PPG signals and compute their average. The final averaged PPG signal is then used as the estimated cardiac signal. And we choose one that is robust to motion and illumination changes [14]. This vision video-based method considers hemoglobin effects on skin appearance along with the effects of illumination changes and selects good local regions to use in cardiac signal extraction. To validate the use of remote video HR estimation for internal emotion detection, we also compare against a wearable sensor. We show that our remote heart rate sensing framework, which uses only a conventional camera, performs favorably in comparison. In addition, our proposed approach has the benefit that commercial cameras are readily available so no special sensors are needed. For example, web cams, surveillance cameras, and cellphone cameras could be used. With the ability to see HR without contact sensors, we present a convenient system for detecting internal emotions.

1.4 Organization of the Thesis

Chapter 2 - Interdisciplinary Background

The computer vision community has made many advancements in emotion recognition through facial expressions. On the other hand, the psychophysiology community has conducted several studies on detecting and recognizing emotions using different physiological channels. Fortunately, in recent years, remote photoplethysmography (PPG) algorithms has received attention. This chapter we talk about cardiac activity from remote PPG to detect and recognize emotions.

Chapter 3 - Detecting Inner Emotions from Video Based Heart Rate Sensing

In Chapter 3, A model is built for inner emotion detection. We use a video PPG method and we chose one that is robust to motion and illumination changes. We experiment to validate the use of remote video HR estimation for inner emotion detection, we proposed an approach commercial camera are readily available so no special sensors are needed. With the ability to see HR without contact sensors, we present a convenient system for detecting inner emotions.

Chapter 4 - Classification of Emotions from Video Based Cardiac Pulse Estimation

In this chapter, we have experiment what kind of video clips we have used. In experimental period all subjects are watching two different kinds of emotional status changes video clips, like comedy and horror video clips. By using their pulse signal, we have analysis and tries to find their emotional status or emotional classification during watching the video clips.

Chapter 5 - Video Method Data set and others feature Algorithm system

To detect emotion recognition, uses of video method is not enough sometime. In this chapter, we have compared HR method with various types of wearable sensors and features, like Wii fit balanced board, pulse oximeter and GSR. Those wearable sensors are used to detect physiological signal.

Chapter 6 – Conclusions

We conclude the thesis with a summary of the concepts and experiment design human inner or real emotion recognition method and human-robot interaction system for human future works and application.

Chapter 2

Related Work

2.1 Emotion recognition from facial expression

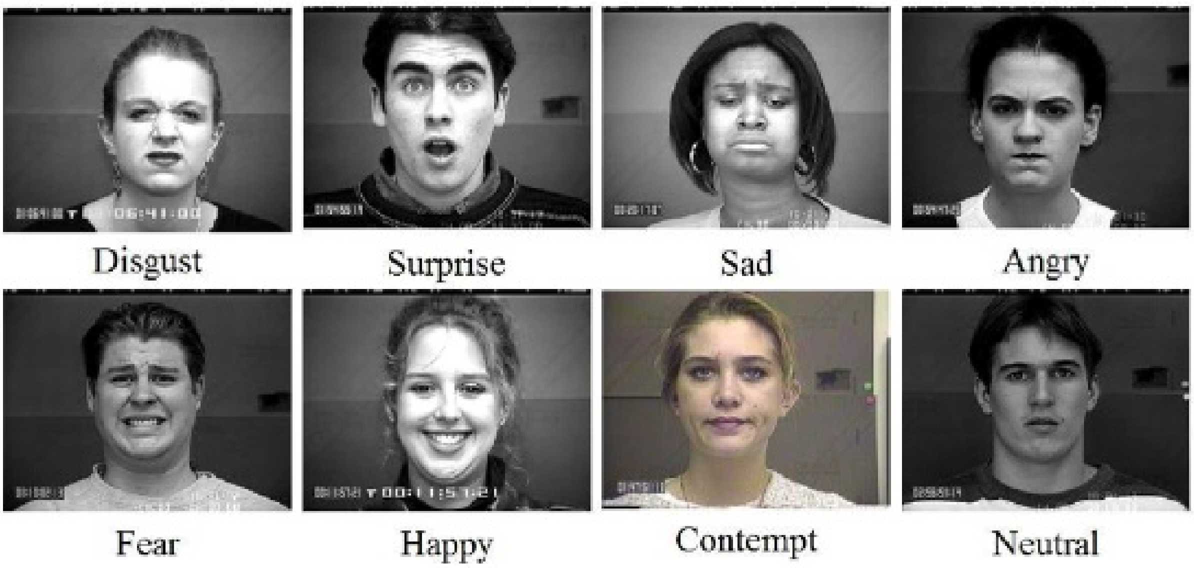


Figure 2.1: Many researchers work on facial expression recognition

A popular approach is to identify emotion from facial expressions. the computer vision community has made many advancements in apparent emotion recognition through facial expressions. Facial expressions give important clues about emotions. Therefore, several approaches have been proposed to classify human effective states. Many researchers work on this field. Though there are methods to identify expressions using machine learning and Artificial Intelligence techniques, this work attempts to use deep learning and image classification method to recognize expressions and classify the expressions according to the images. In contrast the paper focuses to survey

Deep learning techniques used for recognizing the emotions are conducted. Accuracy rate of about 39% is achieved [42].This paper extends the deep Convolution Neural Network (CNN) approach to facial expression recognition task. This task is done by detecting the occurrence of facial Action Units (AUs) as a sub part of Facial Action Coding System (FACS) which represents human emotion. In the CNN fully-connected layers we employ a regularization method called "dropout" that proved to be very effective to reduce overfitting. This research uses the extended Cohn Kanade (CK+) dataset which is collected for facial expression recognition experiment. The system performance gain average accuracy rate of 92.81%. The system has been successfully classified eight basic emotion classes [43]. This paper analyzes the strengths and the limitations of systems based only on facial expressions or acoustic information. It also discusses two approaches used to fuse these two modalities: decision level and feature level integration. Using a database recorded from an actress, four emotions were classified: sadness, anger, happiness, and neutral state. Using markers on her face, detailed facial motions were captured with motion capture, in conjunction with simultaneous speech recordings. The results reveal that the system based on facial expression gave better performance than the system based on just acoustic information for the emotions considered [44]. For a complete review of recent emotion recognition systems based on facial expression the readers are referred to [48]. Masa proposed an emotion recognition system that uses the major directions of specific facial muscles [47]. With 11 windows manually located in the face, the muscle movements were extracted using optical flow. For classification, K-nearest neighbor rule was used, with an accuracy of 80% with four emotions: happiness, anger, disgust and surprise. Yacoob et al. proposed a similar method [50]. Instead of using facial muscle actions, they built a dictionary to convert motions associated with edge of the mouth, eyes and eyebrows, into a linguistic, per frame, mid-level representation. They classified the six basic emotions using a rule-based system with 88% of accuracy. Black et al. used parametric models to extract the shape and movements of the mouse, eye and eyebrows [44]. They also built a mid- and high-level representation of facial actions by using a similar approach employed in [50], with 89% of accuracy. Tian et al. attempted to recognize Actions Units (AU), developed by Ekman and Friesen in 1978 [45], using permanent and transient facial features such as lip, nasolabial furrow and wrinkles [49]. Geometrical models were used to locate the shapes and appearances of these features. They achieved a 96% of accuracy. Essa et al. developed a system that quantified facial movements based on parametric models of independent facial muscle groups [46]. They modeled the face using an optical flow method coupled with geometric, physical and motion-based dynamic models. They generated spatial-temporal templates that were used for emotion recognition. Without considering sadness that was not included in their work, a recognition accuracy rate of 98% was achieved.

2.2 Emotion recognition from other nonverbal behaviors

During everyday communication, especially face-to-face interaction, vocal and visible behaviors are typically coordinated in ways that provide for their mutual performance. When people talk, they also locate their bodies, assume various postures, direct their eyes, perhaps move their hands, altogether behaving in ways that calls human nonverbal behaviors. many researchers, working on using of human nonverbal behaviors.

Figure 2.2: Nonverbal Behaviors

Like Joann M. Montepare [75] explain how to digital technologies provide researchers with new tools for exploring nonverbal components of interpersonal interactions in digital environments. Yang, P et al [53] This paper examines the important role intercultural nonverbal communication competence plays as intercultural responsiveness in the second language learning classroom. And implications for the second language teachers are discussed. Vinciarelli and Mohammadi [54]talk about Nonverbal communication is the main channel through which we experience internal life of others, including their emotions, feelings, moods, social attitudes, etc., for that nonverbal communication can be used as a viable interface between computers and some of the most important aspects of human psychology such as emotions and social attitudes. But there are some cultural differences in nonverbal communication like eye contact, Facial expression, gestures, touch, smell, posture those are expressions are not same all countries. we must realize that we are all different in the way we perceive the world and use this understanding as a guide to our communication with others.

2.3 Emotion recognition from physiological measures

There are times when facial expressions can either be faked or hidden. Thus, other modalities such as physiological responses should be investigated for detecting and recognizing a person's emotions. There are many physiological signals such as brain activity, blood volume pulse, blood pressure, beart rate, skin conductance, skin temperature and respiration. The analysis of physiological signal is a possible approach for emotion recognition [7]. Thus, several types of physiological signals have been used to measure emotions, based on the recordings of electrical signals produced by the brain (EEG), the muscles (EMG) and the heart (ECG). These indications include signals derived from the Autonomic Nervous System (ANS) of the human body fear, for example, increases heartbeat and respiration rate [36]. The Electromyogram (EMG) that measures muscle activity, the Electrocardiogram (EKG or ECG) that measures heart activity, Electrodermal Activity (EDA) that measures electrical conductivity that is in charge of organs such as sweat glands on the skin, the Electrooculogram (EOG) that measures eye movement, and the Electroencephalogram (EEG) [36] .Emotion recognition from physiological signals can help people with impairments in social interaction, communication and developmental disorders. And there have so many wearable sensors or devices using for detect human physiological signal. As figure 2.3 shown in various types of wearable sensors.

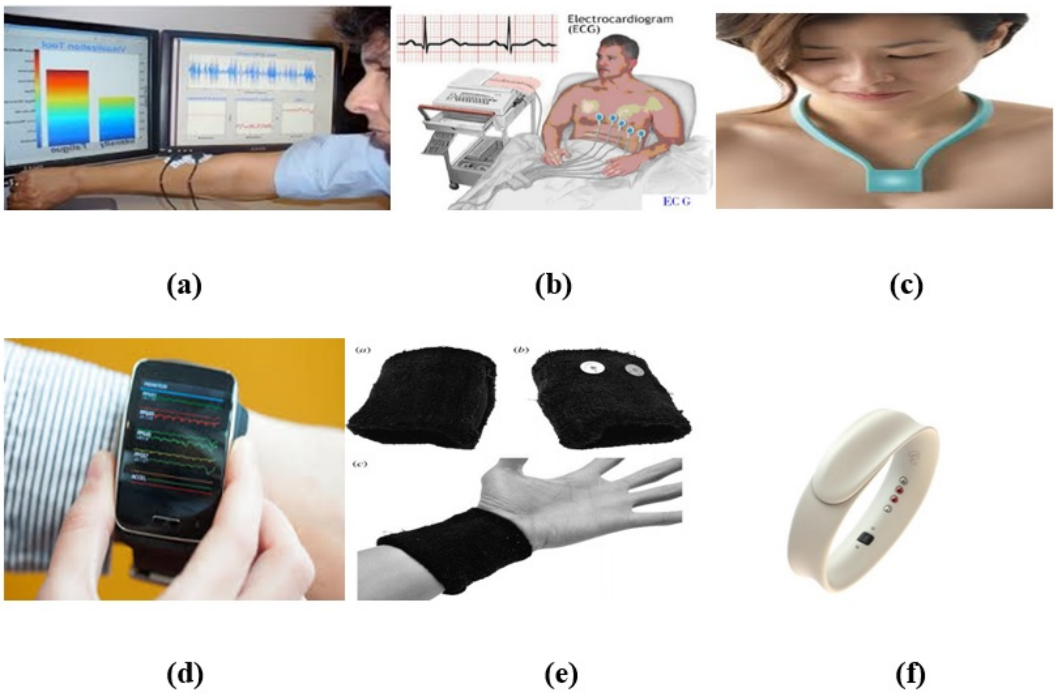


Figure 2.3: There are many types of wearable sensors using for Physiological changes

2.4 Application of emotion recognition: Human Computer Interaction and Human-Robot Interaction

Automatic recognition of emotions in humans is a challenging task with many applications such as human robot interaction, movie marketing, and more. Thus, there has been much work on systems that identify emotional states. Among these areas, the most common application is human computer interaction (HCI). Automatically recognizing human emotion is an interesting and challenging problem. Emotion analysis and recognition has become an interesting topic of research among the computer vision research community, HCI (human computer interaction) and HRI (human robot interaction). One of the interesting challenges in the community of human-computer interaction today is how to make computers be more human-like for intelligent user interfaces. Emotion, one of the users affects, has been recognized as one of the most important ways of people to communicate with each other. Given the importance and potential of the emotions, effective interfaces using the emotion of the human user are gradually more desirable in intelligent user interfaces such as human robot interactions. To make it easier and more natural to interact with robots, people put forward new demands to human robot interaction (HRI). It is hoped that robots can recognize human's facial expressions, understand emotions and give appropriate response. Emotional intelligence robots have attracted great attention in recent years. There are only a few intelligent service systems with emotion. A large body of work in the field of human-robot interaction many researchers have investigated anticipatory robot control through various methods including: monitoring the behaviors of human partners using eye tracking, making inferences about human task intent, and proactive action on the part of the robot. The studies revealed that the anticipatory control helped users perform tasks faster than with reactive control alone. A common approach to program social cues into robots is to first study human-human behaviors and then transfer the learning. These studies have revealed that maintaining a shared representation of the task is crucial for accomplishing tasks in groups [59]. Similarly, researchers have studied the aspect of human-human handovers with household scenarios like passing dining plates in order to enable an adaptive control of the same in human-robot handovers [69]. Most recently, researchers have studied a system that automatically distributes assembly tasks among co-located workers to improve coordination [70]. The application areas of human-robot interaction include robotic technologies that are used by humans for industry, medicine, and companionship, among other purposes. Industrial have been implemented to collaborate with humans to perform industrial manufacturing tasks. While humans have the flexibility and the intelligence to consider different approaches to solve the problem, choose the best option among all choices, and then command robots to perform assigned tasks, robots can be more precise and more consistent in performing repetitive and dangerous work [71]. Emotion detection is also being used in medical applications. A rehabilitation robot is an example of a robot-aided system implemented in health care. This type of robot would aid stroke survivors or individuals with neurological impairment to recover their hand and finger movements[72,73]. In the past few decades, the idea of how human and robot interact with each other is one factor that has been widely considered in the design of rehabilitation robots [73]. Charles Darwin was one of the first scientists to recognize that facial expression is one of the most powerful and immediate means for human beings to communicate their emotions, intentions, and opinions to each other. In addition to providing information about effective state, facial expressions also provide information about cognitive state, such as interest, boredom, confusion, and stress, and conversational signals with information about speech emphasis and syntax. Several ground breaking systems have appeared in the computer vision literature for automatic facial expression recognition

[74.75]. Automated systems will have a tremendous impact on basic research by making facial expression measurement more accessible as a behavioral measure, and by providing data on the dynamics of facial behavior at a resolution that was previously unavailable. Computer systems with this capability have a wide range of applications in basic and applied research areas, including man-machine communication, security, law enforcement, psychiatry, education, and telecommunications.

Figure 2.4: Application of HCI and HRI

Emotion recognition can be used in many sectors: Emotion recognition can be used in Health Care. An industry that's taking advantage of this technology is health care, with AI-powered recognition software helping to decide when patients necessitate medicine or to help physicians determine who to see first. Emotion recognition can be also used in automotive industry. The automotive industry is also applying emotion recognition technology, as car manufacturers around the world are increasingly focusing on making cars more personal and safer for people to drive. The latter is a chiefly attention-grabbing area and one that various companies have by now taken steps in testing and researching. In their pursuit to build more smart car features, it makes sense for car manufacturers to use AI to help them understand the human emotions. Using facial emotion detection smart cars can alert the driver when he is feeling drowsy. Emotion recognition can be used in video game testing. Video games are designed with a specific target audience in mind and aim to evoke a behavior and set of emotions from the users. During the testing phase, users are asked to play the game for a given period and their feedback is incorporated to make the final product. Using facial emotion recognition can aid in understanding which emotions a user is experiencing in real-time as he or she is playing without analyzing the complete video manually [59].

2.5 Chapter Summary

This chapter presented different ways to emotion recognition. It focused emotion recognition systems based on facial expression, various types of nonverbal behaviors, and physiological signals Emotion recognition has possibility of various applications in HCI and HRI.

Chapter 3

Detecting Internal Emotions from Video Based Heart Rate Sensing

3.1 Estimating Heart Rates from Video

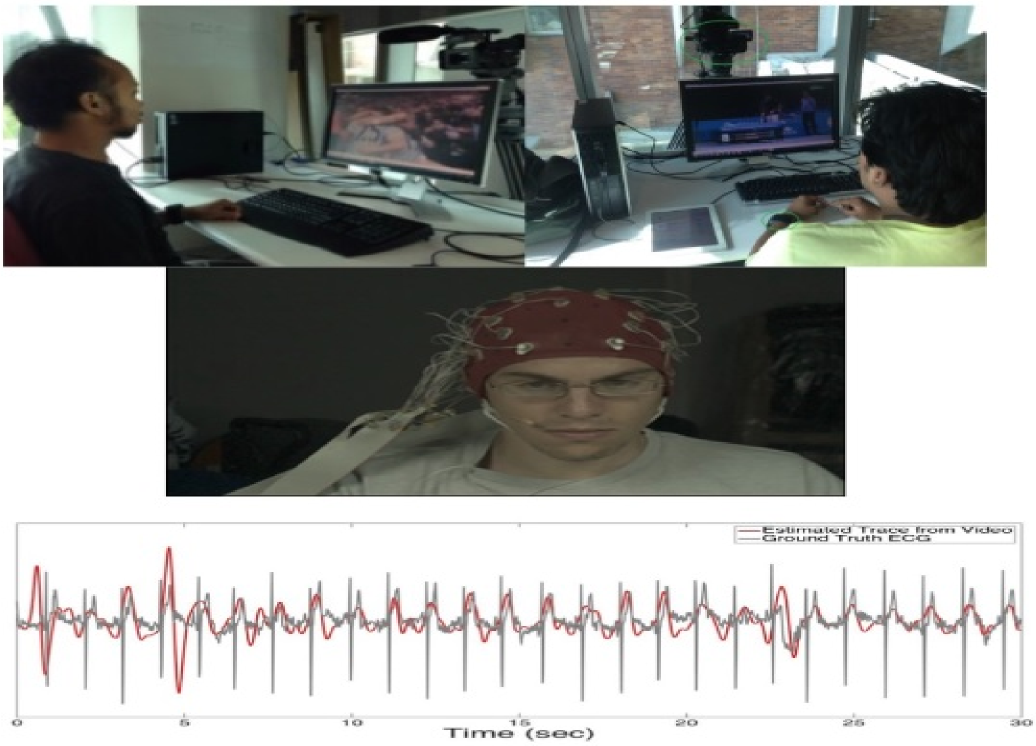


Figure 3.1: Estimating Heart Rates from Video

Human heart rate can be indicative of his/ her stress, fitness, activity level etc. Cardiac pulse is typically measured by electrocardiogram (ECG). But in that way, patients need to wear chest straps with adhesive gel patches. Most of the cases that can be uncomfortable and abrasive for the user. Heart rate may also be monitored using pulse oximetry sensors by wearing fingertip or earlobe. These sensors are not convenient for all the time and need long-term wear and the pressure can become uncomfortable over time. Moreover, discomforts of traditional pulse measurement devices, can damage the fragile skin of premature newborns or elderly people. So, a non-contact or wireless way of detecting pulse could be very beneficial and user friendly. A Non-contact or wireless heart rate measurement through a traditional camera, simple web cam or phone camera would also aid to telemedicine and allow the large number of people to track their heart rate without purchasing special equipment. As part of the recent gain in popularity of fitness apps and the quantified self, regular non-obtrusive monitoring through a computer or phone camera may help detect changes in a person's heart rate over time and indicate changing health or fitness. Heart rate can be detected without contact through photo-plethysmograpy (PPG), which measures variations in blood volume by detecting changes in light reflection or transmission throughout the cardiovascular pulse cycle. PPG is usually performed with dedicated light sources with red or infrared wavelengths, as is the case for pulse oximetry sensors [59]. Verkruysse et al. showed that the plethysmographic signal could also be detected in video from a regular color camera [13]. They found that the signal could be detected within the red, green, and blue channels of color video of exposed skin, but that it was strongest in the green channel, which corresponds to the fact that hemoglobin has absorption peaks for green and yellow light wavelengths. They also found that the signal could be detected in multiple locations on the body, it was strongest on the face, especially on the forehead. Although the plethysmographic signal may be detected in the raw color channel data, it is mixed in with other sources of color variation such as changes in ambient light or motion. Poh et al. found that the signal could be better extracted by using independent component analysis (ICA) to separate independent source signals from the mixed color signals [15]. Other studies have shown that color changes in the face due to pulse may be magnified by amplifying small changes between video frames [11], and that heart rate can be detected through vertical head motion in addition to color changes [23]. Although these are interesting new developments in this space, they are less practical for daily or medical use as the former is more for visualization than quantification, and the latter requires the subject to remain very still for accurate measurements. In this paper, we explore an approach for heart rate detection using RGB color changes in video of faces like that done by Poh et al. [15]. This thesis paper we also using the video camera to recognition estimate pulse signal (Figure 3.1).

3.2 Video PPG Algorithm for Estimating Heart Rate

In recent years, there have been interesting developments in estimating HR from video such as observing very small movements of the head. Balakrishnan et al. [23], exploited subtle head oscillations that accompany the cardiac cycle to extract information about cardiac activity from videos. In addition to providing an unobtrusive way of measuring HR, the method can be used to extract other clinically useful information about cardiac activity, such as the subtle changes in the length of heart beats that are associated with the health of the autonomic nervous system. However, the most common approach is still to consider minute changes in skin color. This approach to reading HR is based on the well-known PPG technique whereby a pulse oximeter contacts skin, illuminates it, and the changes in light absorption due to cardiac pulse is observed. In the case of remote video-based PPG, it has been found that consumer grade cameras (e.g. Canon Power shot) can capture small changes in light absorption (color) that correspond well with the cardiac pulse [13]. Later, this work was extended to account for more realistic imaging conditions [12,14,15,16,17]. These approaches have different pros and cons. In our work, we choose to employ the method by Lam and Kuno [14] because it was designed to address the issue of changing illumination (a problem that can adversely affect color-based HR estimation). This is because our work here involves estimating people's emotional responses to video stimuli on a screen.

For completeness, we briefly summarize the method of Lam and Kuno [14]. The basic model employed assumes skin to consist of two components, hemoglobin, where light absorption is influenced by cardiac activity, and pigments not immediately influenced by cardiac activity such as melanin. Formally, this model expresses the pixel value for a single channel camera of a given point I on skin at time t as

$$I(t) = a_m a_l(t) \int R_m(\lambda) L(\lambda, t) C(\lambda) d\lambda + a_h(t) a_l(t) \int r_h L(\lambda, t) L(\lambda, t) C_k(\lambda) d\lambda \quad (3.1)$$

where $R_m(\lambda)$ $R_h(\lambda)$ are the normalized reflection spectra of the melanin and hemoglobin pigments at wavelength λ,respectively,$L(\lambda, t)$ is the normalized light source spectrum at wavelength λ and time t. Similarly, $C(\lambda)$ is the camera's spectral response. The a terms are constants for scaling the different spectra inside the integration.For example, $a_l(t)$ scales the light spectrum $L(\lambda, t)$ at time t and a $h(t)$ scales

$R_h(\lambda, t)$ at time t. The integration sum over the range of wavelengths λ the camera response $C(\lambda)$ is sensitive to.

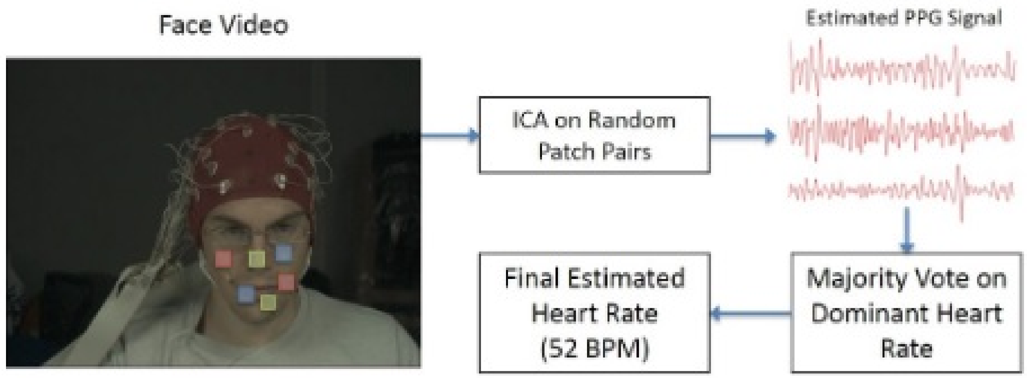


Figure 3.2: Basic Flow of the Video PPG Algorithm for Estimating Heart Rate

We can see from Equation 3.1 that if the light source spectrum (e.g. from a video display) changes colors over time, this affects the amount of light at different wavelengths. This in turn, is reflected by the melanin and hemoglobin pigments differently. Thus, the observed color changes in skin can be thought of as a mixture of two signals. It is well-known that techniques such as Independent Components Analysis (ICA) can be used to perform linear blind source separation (BSS) of signals. Thus, these findings suggest we should be able to take two points on a person's face, obtain the pixel traces of those points, and treating them as signals, apply ICA to separate the melanin and hemoglobin influenced color changes. With the hemoglobin part of the color changes determined, we should be able to calculate the HR. However, our HR method showed that not all pairs of skin surface points could be subjected to ICA for linear BSS. For example, if two skin points were illuminated by different colors, the melanin at the different skin points could reflect very different colors and thus linear BSS would not be applicable. However, provided the two points are illuminated by the same color spectrum (even at different brightness) and the ratios of melanin to hemoglobin between the two skin points are different, ICA can be used to effectively separate out the hemoglobin portion of the signal. Since appropriate skin point pairs are not known a priori, they opted to randomly test using ICA on different pairs of points of the face and observe the histogram of estimated HRs. Then the most common HR in the histogram was used as the final estimate of the HR. The intuition is that randomly chosen point pairs that to satisfy the conditions for linear BSS should consistently give the same HR estimate from the separated hemoglobin portion of the signal. On the other hand, point pairs that violate the conditions for linear BSS should give less consistent HR estimates. As a result, performing a majority vote on the many determined HRs from random point pairs should give a robust estimate for a single final HR. See Figure 1.2 for the basic flow of the algorithm.With a robust algorithm for estimating HR in the presence of illumination changes (referred to as the Vision Based Video Method, VBVM), we proceed to determine whether it can be used for determining the emotional states of people watching real-world videos. This is done in comparison to a commonly available wearable sensor, namely the Fitbit.

3.3 Detect Internal Emotion from Video Vision Based

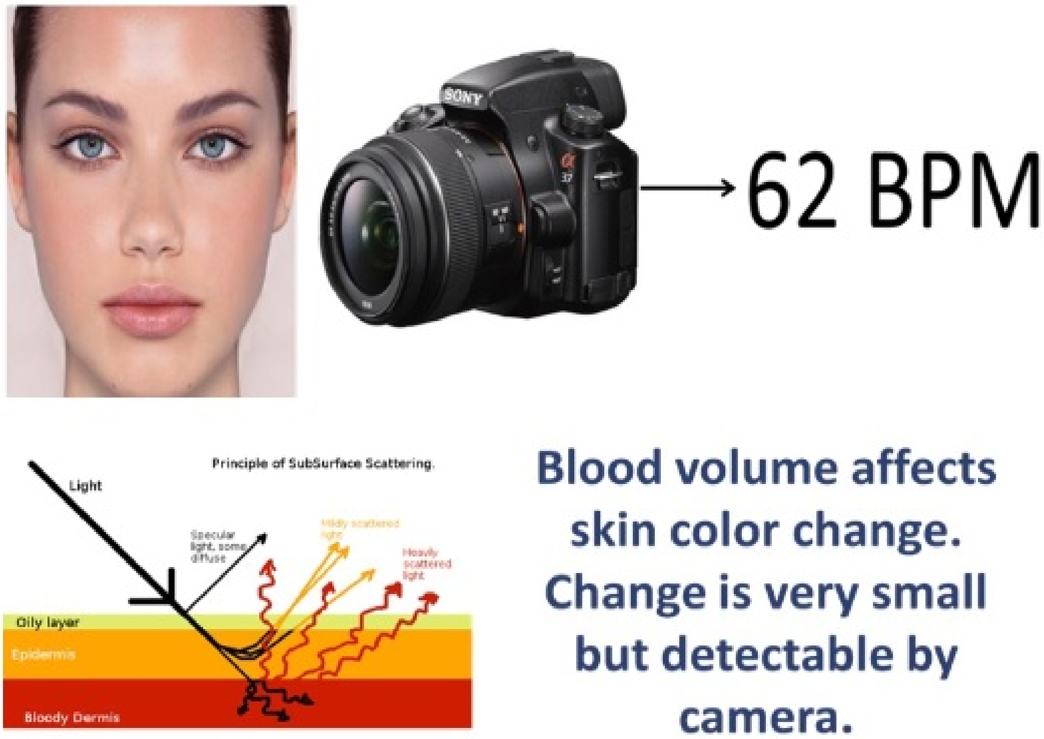


Figure 3.3: Using Video Camera for Emotion detection

Now a days, it is very popular to detect human's facial expression and emotion recognition from videos cameras. Using camera detect face very popular now (Fig 3.3). Many researchers already work on this field. Many real time applications based on human computer interaction systems are used to track the human activities from the videos. This system is realizing and tracking the human face expression and emotions recognition from the video streaming with an objective of different. However, before going to discuss about it more precisely, first we introduce about needs of facial emotion and expression recognition.

In human-to-human conversation, the sound of mental, emotional, and even physical state is used in conversations about important information in addition to pronounce a communication channel and facial expressions is the notion of a persons facial expressions in its simplest form is a more subtle happy or angry thoughts [61], feelings or absorption of all speaker expectation from listeners, sympathy, or even what the speaker is saying no signal can provide to computing background, bring our everyday human user to remain at the forefront in the fabric will move to absorb that predict a generally establishment [62]. Recognition of emotions from facial expressions using videos consists of preprocessing, feature pulling out and division. Importance of facial expression system is widely recognized in social interaction and social intelligence system analysis is an active research topic since the 19th century. Suwa ET was introduced facial expression recognition in 1978 Al. creating a facial expression recognition system the main point of face detection and alignment Feature extraction and classification, image standardization [63]. Using this method can identify facial expressions match there to recognize that there are four types of expression [64]. Facial expression to recognize the first type uses emotion s speed. The second type of optical flow using facial expression to identify an image frame is the third Type facial expression to recognize the active shape model to use. The fourth type neural networks [65] using facial expression to recognize face a complex multidimensional view. Model and to develop a model for face recognition is hard work. Face detection there are available several types of different condition database (expression, Lights, etc.) with a different face [66].

Emotion recognition has become an important and interesting field of study in Human- Computer Interaction (HCI), Human Robot Interaction (HRI), etc. The six basic emotions are, sickening, happy, fear, anger sad and surprise. Computer graphics, automatic driver fatigue detection, 3D or 4D avatars animation in the entertainment industries, psychology, video & amp; text chat and gaming applications are include in diverse applications.[60]. However, there are very less concentration on inner emotion or real emotions recognition. as we know that human face always not show their real expression. Sometimes human can be able to hide their emotions or show fake emotions. So that it was very important to take the attention about the human inner emotion detection. In this chapter we work on detection human inner emotions using only video camera. We proposed by using our vision-based video method to estimate cardiac pulse signals and recognize emotion from the pulses. We show that from the cardiac pulses itself, we can be able to determine the type of video (comedy vs. horror) while the human (subject) watching. As shown (Fig 3.3) when subject watching video his facial expression remains same but by analyzing their pulse signal, we find their emotional status has changed. Next experimental part we discuss in details.

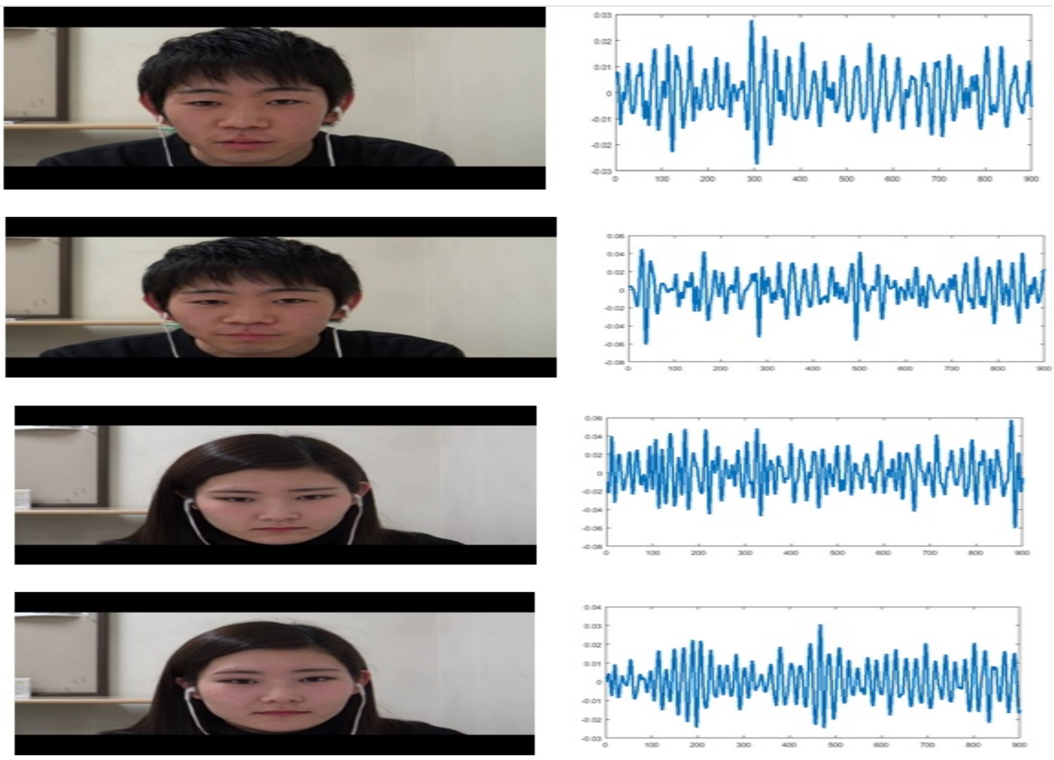


Figure 3.4: Getting cardiac pulse signal

3.4 Experimental Analysis and Result

For this study, we make an experiment flowchart (Fig 3.4). 40 undergraduate and postgraduate engineering students from Saitama University were recruited for data collection. The participant's ages ranged from 20 to 25 years old and there were 25 males and 15 females. They were all Asian descent and in sound health. The experiment consisted of three parts. In the first part, we measured the human subject's resting HR for 1 minute. Then using a Sony 4K FDR- AX30 camera, we recorded 29 FPS videos of each subject's face for 4 minutes while the subject watched a funny video. After that, we recorded the subject for 4 minutes while he/she watched a

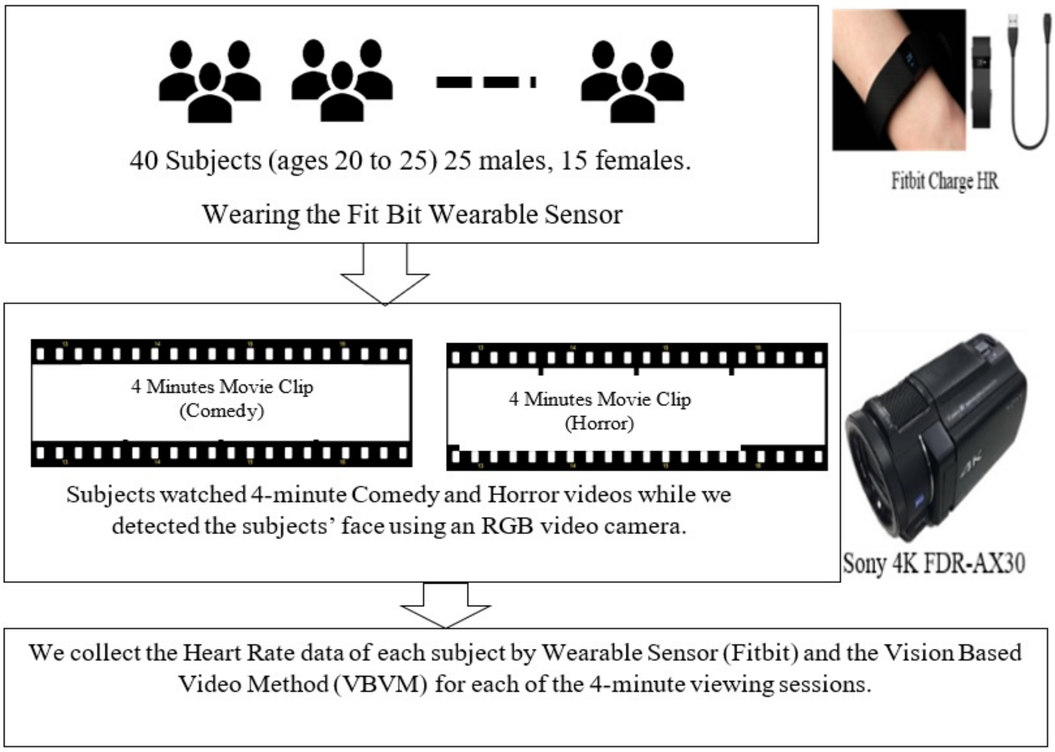


Figure 3.5: Experiment design

horror video clip. In all cases, we determined the average HR for each minute of the viewing sessions.

We should briefly discuss our choice in using the Fitbit sensor (Fig 3.5). Concerning commercial wearable sensors, there are many on the market. Generally, companies display advertising for these kinds of products and depict them as beneficial, user friendly, and accurate but it is important of consider the research literature to objectively determine their usefulness. Kaewkannate et al. [18] in their paper summarizes and compares wearable fitness devices. In addition, Evenson et al. [19] proposed summarizing the evidence for validity and reliability of the most popular consumer-wearable activity trackers. Among a variety of trackers on the market, approximately 3.3 million were sold between April 2013 to March 2014, with 96% of the sensors made by Fitbit (67%), Jawbone (18%), and Nike (11%) [19]. In their meta-study, they reported that the Fitbit was often found to have high reliability for monitoring steps, energy expenditure, and sleep in some Fitbit models but that none were reported for the Jawbone. Due to the better understood performance characteristics of the Fitbit for various monitoring tasks, we also chose to use the Fitbit. In addition, internal studies by Fitbit report their trackers are 95-97% accurate.

Figure 3.6: Experiment's setup for measuring heart rate during watching video clips.

3.4.1 Result for Single Subject

Here we show an analysis of Vision Based Video Method (VBVM) HR estimation and wearable sensor HR from funny and horror video clips. We compare the HR data from VBVM and the wearable sensor (Fitbit). We found that the two methods give us similar results from two emotional situations. In fact, the two methods had HR estimates with a correlation of 0.90 (Figure 3.6 and Figure 3.7) also shows the mean HRs (over the entire viewing session for each subject) for the VBVM and wearable sensor are consistent across different subjects. Thus, the VBVM provides comparable performance to the wearable sensor.

We collected all subjects' HR data using VBVM and the wearable sensor (Fitbit). We observed them while they watched video clips (funny and horror) and compared their resting HRs with their HRs while viewing the videos (taken as the maximum between the two to three-minute points in each viewing session because these were the most emotionally stimulating parts). In addition, Figure 3.8 shows t-test results for various cases. These tests compare the HRs during emotional stimulation relative to the mean resting HR (69 ± 5.78 BPM) of the subjects. All the p-values of these

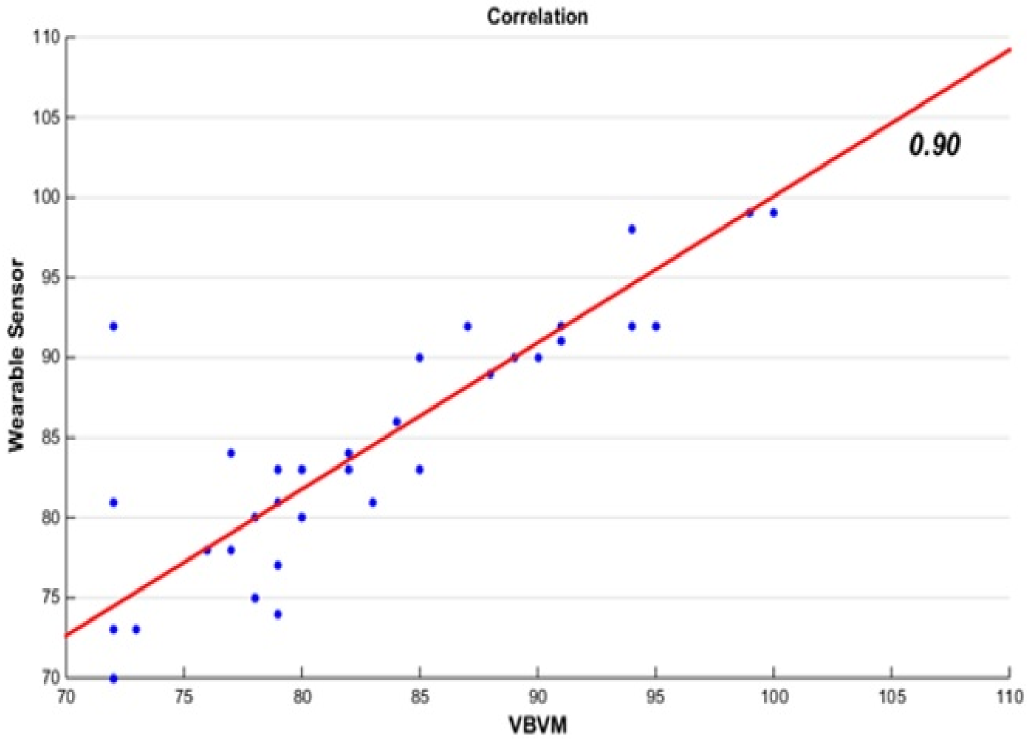


Figure 3.7: Correlation of the VBVM and Wearable sensor

tests show clear statistical significance indicating that when emotionally stimulated, people do experience increased HRs. Both the VBVM and wearable sensor pick up these changes in HRs reliably.

Fig 3.8 shows the bar graphs of the resting, funny, and horror cases for both VBVM and the wearable sensor data. From the diagram, we see that the HR increases by about 22%-26% and for the funny video case and 18%-19% for the horror video case. The HR for the funny video case is typically higher than the horror video case for both VBVM and the wearable sensor.

3.4.2 Result for Multiple Subjects

We now present a demonstration of our system measuring the HR of multiple human subjects in our lab. In this demonstration, we used a funny video clip in English. We selected three subjects for the demonstration. We then recorded the three subjects watching the video clip. The video was then used to estimate their HRs over 12 minutes and the plot of these HRs can be seen in. The average HRs for subjects A, B, and C were estimated at 74, 77, and 70 respectively, which is within the normal range of HRs. We can also see from Figure 3.6, that in some parts of the viewing

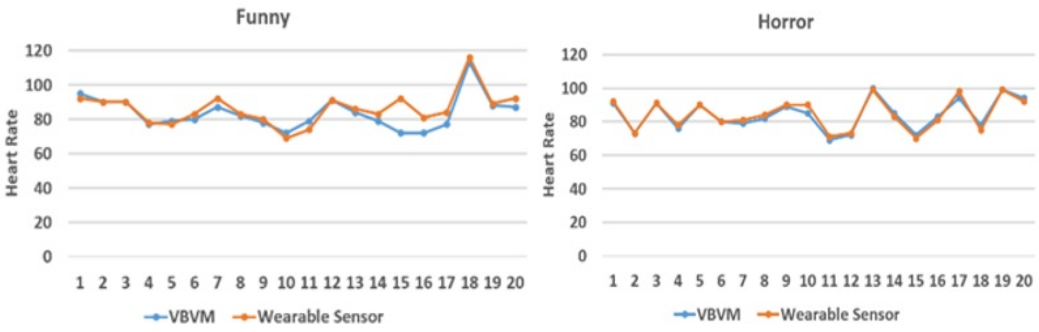


Figure 3.8: Sample mean heart rates for 20 subjects from the viewing sessions.

session, the HRs increased and decreased at the same times. However, since these subjects had varying levels of English comprehension, we believe there were some differences in their appreciation of the humor, which would result in some differences in HR changes over the session between the subjects. In future work, we will consider videos in languages more suited to the human subject being tested.

We also note that Subject B (the middle person in Figure 3.9) did not show much change in facial expressions. However, we see from Figure 3.10, that his HR had the most variability. Subject A and Subject C showed smiling faces but their HRs had less changes. From this demonstration, we can see an example of how facial expressions may not always indicate genuine emotions. But physiological responses such as HR can reveal a person's hidden emotional responses.

3.5 Conclusion

In this chapter, we collected HR data for three emotional states (normal resting, funny, and horror) and performed a statistical analysis of the results. The HR data received using the VBVM is strongly correlated with the wearable sensor ground truth data. Moreover, our experiments results show that the HRs for the funny clips increase approximately 24.04% and there is a 19.06% rise for horror clips. We were able to detect emotional state changes from HR. Our method has some limitations like extreme changes in illumination and rapid motion that affect accuracy. In the future, we will resolve these problems to improve our system performance. Another interesting line of future work is that we will develop an approach to automatically decide whether a person has experienced inner emotional change. The current work here establishes the ground work by statistically verifying the feasibility of the basic sensors used in our framework but we currently have no automated way to decide if

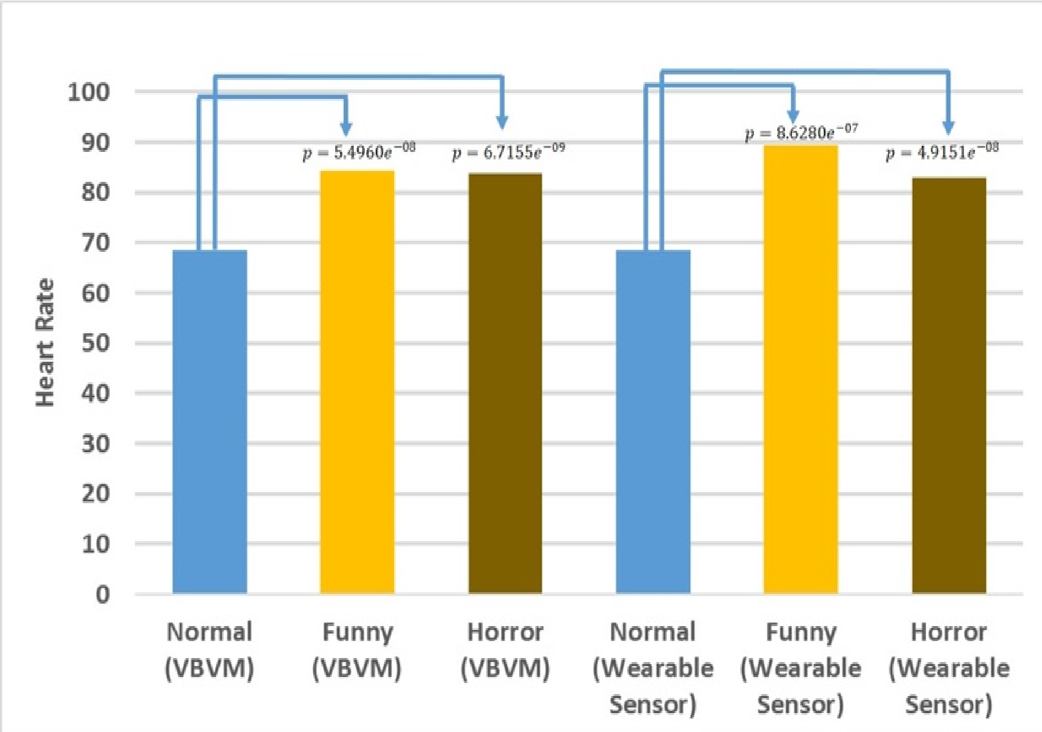
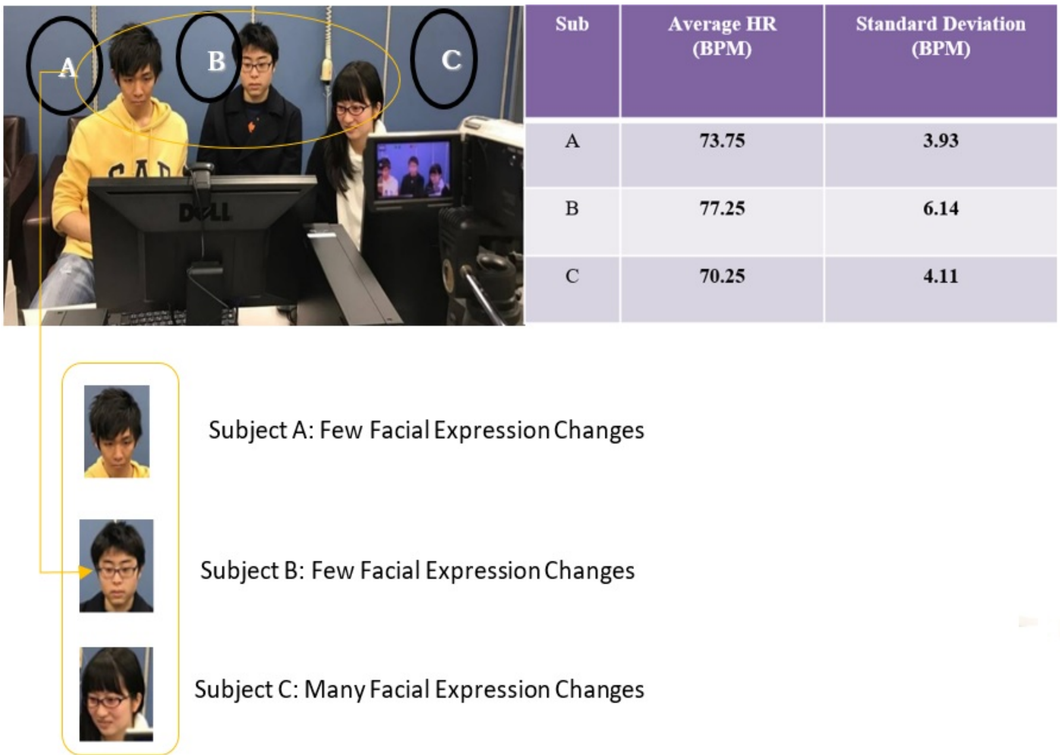


Figure 3.9: Average HR bar graphs of the Resting, Funny, and Horror cases.

someone is reacting to emotional stimuli. (For example, a Support Vector Machine or other machine learning technique could be used to take in various HR data and classify people's reactions.) Moreover, we will account for additional emotional states like sadness, anger, and frustration. Differentiating between the various types of emotions will likely require some form of detailed subtle facial expression analysis but detecting changes in HR would still be needed to indicate the presence of emotional stimulation as a first step. We also plan to continue our work to evaluating multiple people at the same time (as is done in the demonstration at the end of the paper). This would be particularly interesting as we could then observe the emotional reactions of audience members in movie screenings. This would be useful for pilot screenings so that movie producers could adjust content to improve the quality of movies. An interesting idea might even be to have a system that could adaptability alter the movie's content based on the sensed emotions in order to enhance the emotional experience of watching a movie. In addition, we will also investigate using the cardiac PPG signal itself and computing metrics such as HR variability, which is known to also be a good indicator of emotional change. It would be interesting to see if observing the continuous changes in the cardiac PPG signal itself could provide even more information about a person's



Sub	Average HR (BPM)	Standard Deviation (BPM)
A	73.75	3.93
B	77.25	6.14
C	70.25	4.11

Figure 3.10: Measuring heart rate during viewing of video clips for three different subjects. From left to right, subjects A, B, and C.

emotional state.

3.6 Chapter Summery

This chapter we presented measurement heart rate from video. we use a remote video-based heart rate sensing technique to obtain physiological data that provides an indication of a person's inner emotions. This method allows for contact-less estimates of heart rate data while the subject is watching emotionally stimulating video clips. We also compare against a wearable heart rate sensor to validate the usefulness of the proposed remote heart rate reading framework. We found that the reading of heart rates of a subject effectively detects the inner emotional reactions of human subjects while they were watching funny and horror videos—despite little to no facial expressions at times. This chapter show that maybe it possible we can be use HR from video can be used to estimate inner emotions.

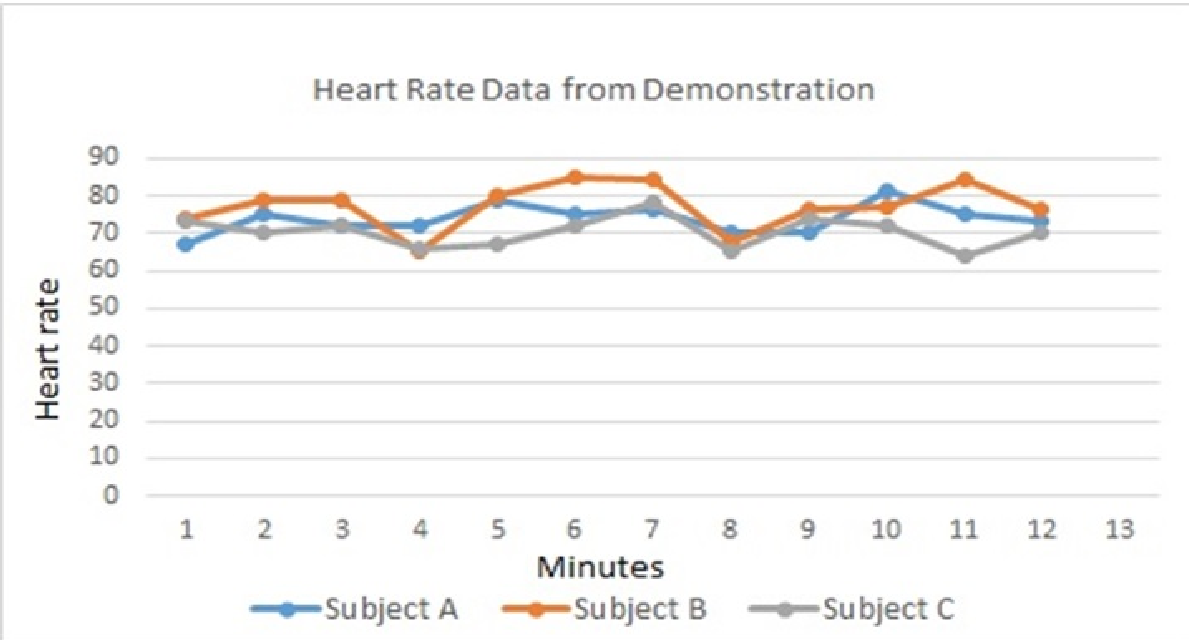


Figure 3.11: Comparison between Heart Rate Data for Three Subjects

Chapter 4

Classification of Emotions from Video Based Cardiac Pulse Estimation

In this chapter, we propose using our vision-based video method to estimate cardiac pulse signals (as opposed to only the heart rate) and recognize emotion from the pulses. We show that from the cardiac pulses alone, we can determine the type of video (comedy vs. horror) the human subjects are watching. We also compare our approach with a method using facial action units for emotion recognition.

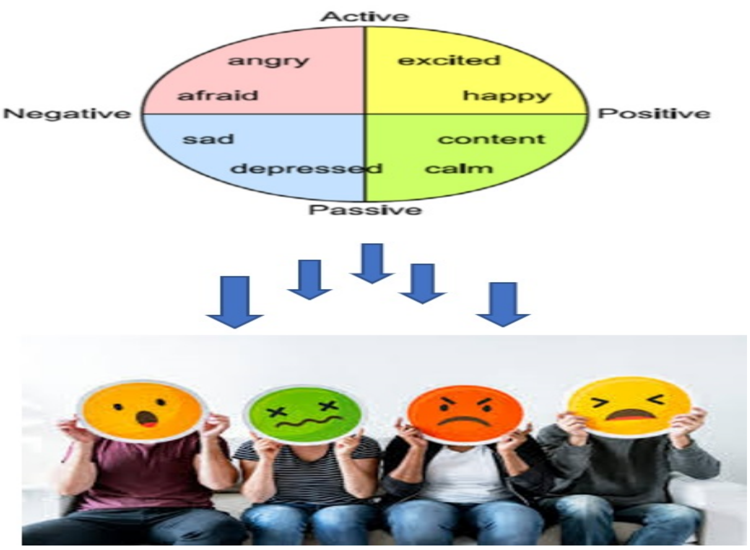


Figure 4.1: Different types of emotion

We modify the heart rate measurement method so that it can estimate heart pulse signals. Then, we develop a recognition method using support vector machine

(SVM) and principal component analysis (PCA). we conduct an experimental study a recognition emotion when people watching video clips.After watching different types of movie there inner feeling or emotion possibility to changes but looking their facial expression sometimes is very difficult know their emotion or feeling change or not. In this part we try to classifying their emotions when they watching video clips. For this purpose, this part we using support vector machine (SVM) and Principal Component Analysis (PCA) to classifying inner emotions.

4.1 Estimating Heart Pulse from Video

The method described in section 3.2, extracts heart pulse signals by using ICA and then estimate the heart rate from them. We cannot recognize emotion from heart rate alone. We use the heart pulse signals for this purpose.to do this, we need more precise signal estimation method. In the original method, we use small square patches. Such square patches may not be always on the same face surface positions because the face may move. Thus, we detect facial landmarks obtained by face indicate the method, reference to detect landmarks and use Delaunay Triangulation using these landmarks. We measure the average intensity in each triangle and use the data for ICA. In addition, like Lam et al. [27], we also employ the RGB channel weighting scheme proposed by Haan and Jeanne [28]. Here our goal is to use the cardiac pulse signal (PPG signal) to perform emotion recognition. So, after estimating the final HR, we go back and identify all the PPG signals that contributed to the final HR estimate. We then normalized each of those PPG signals and compute their average. The final averaged PPG signal is then used as the estimated cardiac signal.

4.2 Classification Method

In this work, we try to understand the emotional information in facial action units and pulse signals for the emotional states (happy and scary) using our method. In works related to emotion recognition using physiological signals, acquiring emotional physiological data is challenging because of the subjective nature of emotions and cognitive dependence of physiological signals. Researchers have used different methods to elicit the target emotions such as visually showing images, using audio media such as music, and audio-visual stimuli in the form of short film video clips. In this work, emotions were induced by using short video clips. Twenty-six subjects in the aged 20-25 years old participated in our study to rate the emotions they experienced while watching the video clips. They were all Asian descent and in sound health. We recorded videos of each subject's face for a total of 10 minutes while the subject watched movie video clips. In each video segment, the first 5 minutes, they watched a horror video clip. A wearable heart rate sensor (Fitbit) was also used for simple verification of their general cardiac responses. After collecting our videos, we chose 30 seconds of the videos where the peak emotional responses would be expected. For example, in the case of comedy, we chose the 30 second segment of the video with the punchline for the joke. In the case of horror, we chose the 30 second segment with the scariest part of the video. Figure 4.4 shows sample screen shots from the videos of the subjects and their estimated cardiac pulses [23]. We also used Open Face to obtain intensities for 17 facial action units for each frame in the videos we analyzed. These facial action units were obtained to use as a means of comparison between our cardiac based approach and a facial expression-based approach. For each 30 second clip, we determined the average intensity of each facial action unit resulting in 17 values associated with each 30 second video segment to be used for emotion recognition.

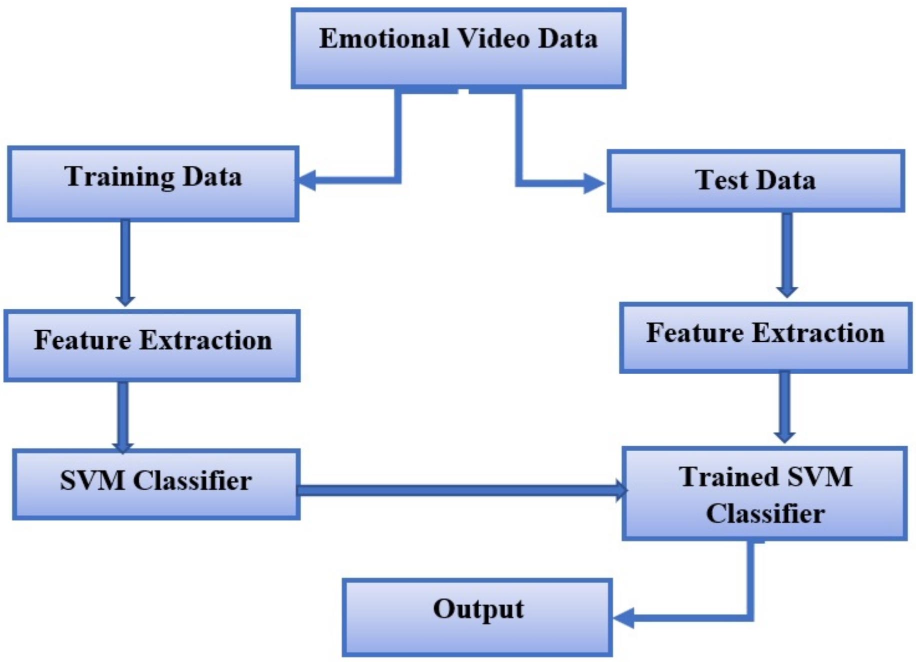


Figure 4.2: Diagram of classification inner emotion

4.3 Using Support Vector Machine (SVM) and Principal component analysis (PCA) for Classification

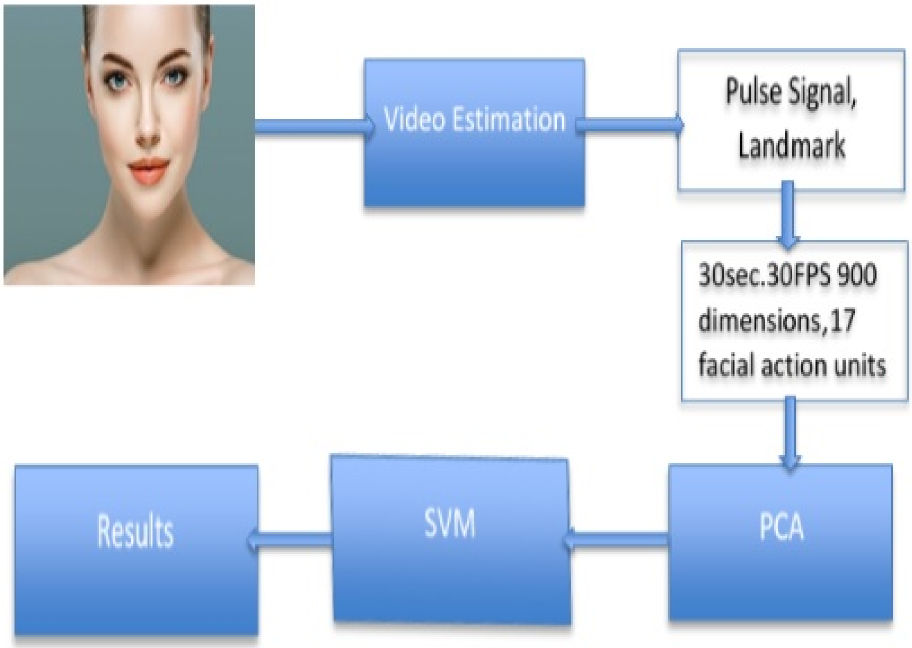


Figure 4.3: For Classification use the SVM and PCA

4.3.1 Using Support Vector Machine (SVM) for emotion classification

We use a support vector machine to recognize emotion on the ICA compressed data. A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyper plane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyper plane which categorizes new examples. In two-dimensional space this hyper plane is a line dividing a plane in two parts where in each class lay in either side. We present results on emotion classification using remotely estimated cardiac pulse signals from RGB video. We then used a linear SVM to train a model based on ground truth labels of the video clip's genre. In order to evaluate generalization performance, we performed leave-one-out cross-validation on our dataset, which consists of 26 instances of reactions to horror and 26 reactions to comedy. Thus, in each leave-one-out cross-validation test, 51 instances were used for training and 1 instance was used for testing. The average accuracy was then determined.

4.3.2 Principal component analysis (PCA)

Principal component analysis (PCA) is a technique used for identification of a smaller number of uncorrelated variables known as principal components from a larger set of data. The technique is widely used to emphasize variation and capture strong patterns in a data set. principal component analysis is a tool used in predictive models and exploratory data analysis according to the Wikipedia. Principal component analysis is considered a useful statistical method and used in fields such as image compression, face recognition, neuroscience and computer graphics. We also compare against using 17 facial action units to classify emotional response. In order to classify emotional responses, we took each of the 900-dimensional cardiac pulse signals and treated them as vectors. For the case of facial action units, we also performed the same tests. In addition, we decided to test preprocessing our data using PCA (mainly because of the high dimensionality of the cardiac pulse signals). We chose to project both the cardiac pulse signals and facial action units onto their respective eigenvectors such that 90% of the variance would be retained.

4.3.3 Using Support Vector Machine (SVM) for emotion classification

We use a support vector machine to recognize emotion on the ICA compressed data. A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two-dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side. We present results on emotion classification using remotely estimated cardiac pulse signals from RGB video. We then used a linear SVM to train a model based on ground truth labels of the video clip's genre.

4.4 Experiment

In this section, we present results on emotion classification using remotely estimated cardiac pulse signals from RGB video. In each case, the video viewing sessions lasted about 30 seconds and the videos were captured at about 30 FPS. This would imply

Figure 4.4: Find the different kind of pulse signal for same facial expression

that each cardiac pulse signal would result in 900 dimensions as there would be 900 frames per viewing session. Since the timings were not exact, some of the estimated cardiac pulse signals had slightly more than 900 dimensions. For example, 905 dimensions. We chose to simply discard the small number of extra dimensions so that all cardiac pulse signals would consist of exactly 900 dimensions. We also compare against using 17 facial action units to classify emotional response. Our classification task is to determine which genre of video the subject watched. In the case of the horror clip, we expect most people would experience fear and in the case of the comedy clip, we expect people would experience a combination of joy and happiness.

Thus, if we can classify the genre of video the human subject watched, we essentially classify their emotional reaction. In order to classify emotional responses, we took each of the 900-dimensional cardiac pulse signals and treated them as vectors. We then used a linear SVM to train a model based on ground truth labels of the video clip's genre. In order to evaluate generalization performance, we performed leave-one-out cross-validation on our dataset, which consists of 26 instances of reactions to horror and 26 reactions to comedy. Thus, in each leave-one-out cross-validation test, 51 instances were used for training and 1 instance was used for testing. The average accuracy was then determined. For the case of facial action units, we also performed the same tests. In addition, we decided to test preprocessing our data using PCA (mainly because of the high dimensionality of the cardiac pulse signals). We chose to project both the cardiac pulse signals and facial action units onto their respective eigen vectors such that 90% of the variance would be retained. The results can be seen in Table 4.1.

Table 4. 1 Leave One Out Cross-Validation Results. We used linear SVMs to learn the classifiers in all cases. The term Video PPG denotes the cases where remotely estimated cardiac pulse signals from RGB video were used as the features for learning. Without using PCA to reduce dimensionality, we see that video PPG results are negatively impacted. With PCA, the results are not as good as in the case with facial action units but our proposed approach has the advantage that not relying on facial features, the emotional responses cannot be as easily faked.

Table 4.1: We used linear SVMs to learn the classifiers in all cases

Feature Type	Leave One Out Cross-Validation Accuracy
Linear SVM with Video PPG	65.4%
Linear SVM with Video PPG (PCA)	67.3%
Linear SVM with Facial Action Units	76.9%
Linear SVM with Facial Action Units (PCA)	78.8%

In Table 4.1, we see from the “Video PPG” results that emotional reactions can be classified from remotely sensed cardiac activity alone. Furthermore, the cardiac pulse signals were estimated from a conventional RGB camera so no special equipment was required. From the first row’s result labeled, “Linear SVM with Video PPG”, we see that the result gives a 65.4% accuracy. Since the dimensionality of the cardiac pulses was high (900 dimensions), we also decided to test using PCA to reduce the dimensionality of the cardiac pulse signals. We chose to project the cardiac pulse data onto eigen vectors such that 90% of the variance would be retained and then test again using Linear SVMs. We found that doing so did increase the accuracy to 67.3%. In the case of facial action units, we used 17 features and found that the leave one out cross-validation accuracy without PCA resulted in a 76.9% accuracy and with PCA, 78.8%. These classification results are better than our proposed cardiac-based classification approach. However, we believe this is because we did not instruct the participants to hide their genuine emotions. As a result, our dataset was unfortunately not ideal for illustrating the effects of faked emotions on facial action unit-based emotion recognition. Despite this setback, we did find that our "Video PPG" based approach is able to recognize emotional reactions with reasonably good performance. The major advantage of our approach is that since we do not rely on facial expressions, the emotional reactions we detect cannot be easily faked. This is because our sensing is based on physiological reactions.

4.5 Conclusion

We have presented an approach to emotion recognition that is based on physiological responses rather than facial expressions. As a result, the emotions we detect cannot be easily faked. In addition, these cardiac pulse signals were entirely estimated from videos captured by a conventional RGB camera and so no special equipment is required. Essentially, we can realize a system that operates completely using only computer vision techniques. we conducted an experimental study on recognizing the emotions of people watching video clips. We recorded every subject; all watched the same comedy and horror video clips and then we estimated their cardiac pulse signals from the video footage. From the cardiac pulse signal alone, we were able to classify whether the subjects were watching the comedy or horror video clip. We also compare against classifying for the same task using facial action units and discuss how the two modalities compare. A drawback of the current study is that our dataset does not have human subjects that intentionally tried to fake their emotions and so it was not ideal for our tests. However, we were able to show that even with a naive approach like taking the estimated cardiac pulse signals, performing dimensionality reduction using PCA, and then leaning via linear SVM, we achieved surprisingly good accuracy. In the future, we will continue to investigate better ways to extract the relevant features from the cardiac pulse signals to improve accuracy. Given the fact that our system only requires a conventional RGB camera, it would be interesting to explore various applications in effective computing in the future.

Table 4.2: These are some dada set video clips subjects emotions

Clip Id	Emotion	Clip Names	Video Clip Sources
1	FL1	Avi.003	Funny baby Horror Clips (YouTube)
2	FL1	Avi.050	Funny baby Horror Clips (YouTube)
3	FL1	Avi.004	Funny baby Horror Clips (YouTube)
4	FL1	Avi.070	Funny baby Horror Clips (YouTube)
5	FL1	Avi.035	Funny baby Horror Clips (YouTube)
6	FL1	Avi.041	Funny baby Horror Clips (YouTube)
7	FL1	Avi.005	Funny baby Horror Clips (YouTube)

4.6 Chapter Summery

We modify the heart rate measurement method using video so that we can obtain heart pulse signals. Then we develop an emotion classification method from the heart pulse signals. This chapter shows a possibility of recognizing human emotions from video.

Chapter 5

Hidden Emotion Detection Video method and others Feature comparison

Emotion recognition is an important research topic. Physiological signals seem to be way for emotion recognition and now a days to possible that wearable sensors are required to collect these data. Therefore, wearable sensors are commonly used while the number of wearable devices including similar physiological sensors is growing up. Many studies have been completed to evaluate the signal quality obtained by these sensors but without focusing on their emotion recognition capabilities. In the study, this thesis we compare video base pulse signal for human emotion recognition data and various wearable sensor in terms of emotion recognition accuracy. This Thesis paper we already talk about that we use a remote video-based heart rate sensing technique to obtain physiological data that provides an indication of a person's inner emotions.To validate the use of remote video HR estimation for inner emotion detection, we also compare one against a wearable sensor(Fitbit).With the ability to see HR without contact sensors, we present a convenient system for detecting inner emotions. But this time we compare various types of wearable sensors or feature for our method.

5.1 Introduction

Figure 5.1 shows the wearable sensor this part we use for compare our method.in many researchers use of wearable sensor for their studies, this thesis paper we use those of sensor only for compare. Our main intention is investigating pulse signal related to emotions and identify the intrinsic patterns of pulse signal for three emotional states: excited, relax and boring. We examine various modality fusion strategies for

(a) Pulse Oximeter

(b) Wii fit balanced board

(c) Shimmer GSR

Figure 5.1: Various types of wearable sensor or feature

integrating users' external subconscious behaviors and internal cognitive states and reveal that the characteristics of pulse signal and wearable sensor are complementary to emotion recognition. Pulse oximetry method for monitoring a person's oxygen saturation (SO2). In its most common application mode, a sensor device is placed on a thin part of the patient's body, usually a fingertip, or in the case of an infant, across a foot. Reflection pulse oximetry is a less common alternative to transmit pulse oximetry. This method does not require a thin section of the person's body and is therefore well suited to a universal application such as the feet, forehead, and chest, but it also has some limitations. Vasodilation and pooling of venous blood in the head due to compromised venous return to the heart can cause a combination of arterial and venous pulsations in the forehead region and lead to spurious SpO2 results. The Wii Balance Board is shaped like a household body scale, with a plain white top and light gray bottom. Wii Balance Board, with four directional switches instead of pressure sensors. shift of center of pressure (CoP) is an indirect measure of postural sway and thus a measure of a person's ability to maintain balance. Though originally designed as a video game controller, the Balance Board has become a tool for assessing CoP which has proven to be both valid and reliable. Clark et al.[40] performed a study to prove the validity and test-retest reliability of the use of a Balance Board. The reason to use a Balance Board instead of a force platform is the ability to "create a portable, inexpensive balance assessment system that has widespread availability." Four standing balance tasks were used in this study including a combination of double stance, single stance, eyes open, and eyes closed. Throughout these tests the center of pressure path length was measured and compared to data from an identical study on a laboratory-grade force platform. The study found Balance Board measurements to be reliable and consistently repeatable. Those of feature we compare our method. Shimmer3 GSR+ (Galvanic Skin Response) unit provides connections and preamplification for one channel of Galvanic Skin Response data acquisition (Electrodermal Resistance Measurement - EDR/Electrodermal Activity (EDA). The GSR+ unit is suitable for measuring the electrical characteristics or conductance of your skin, as well as capturing an Optical Pulse/PPG (Photoplethysmogram) signal and converting to estimate heart rate (HR), using the Shimmer ear clip or optical pulse probe. The Galvanic Skin Response Sensor is used for real time GSR Biofeedback. The Shimmer GSR+ sensor monitors skin conductivity between two reusable electrodes attached to two fingers of one hand caused by a stimulus the sweat glands become more active, increasing moisture on the skin and allowing the current to flow more readily by changing the balance of positive and negative ions in the secreted fluid (increasing skin conductance).

5.2 System Overview

Emotion recognition based on physiological signals has been a hot topic and applied in many areas such as safe driving, health care and social security. In this thesis paper, we present a comprehensive review on physiological signal-based emotion recognition, including emotion models, emotion elicitation methods, the published emotional physiological datasets, features, classifiers, and the whole framework for emotion recognition based on the physiological signals. A summary and comparative among the recent studies have been conducted, which reveals the current existing problems and the future work has been discussed.

In this thesis paper, we adopt a multimodal emotion recognition framework by combining video method heart rate and electroencephalography (EEG) to enhance emotion recognition. The main contributions of this thesis paper propose heart rate variability (HRV) features extracted from photoplethysmogram (PPG) signal obtained from a cost-effective PPG device such as Pulse Oximeter for detecting and recognizing the emotions based on the physiological signals. The HRV features obtained from both time and frequency domain are used as features for classification of emotions. These features are extracted from the entire PPG signal obtained during emotion elicitation and baseline neutral phase. For analyzing emotion recognition, using the proposed HRV features, standard video stimuli are used. We have considered three emotions namely, excited, relax and boring emotions. Detection of true

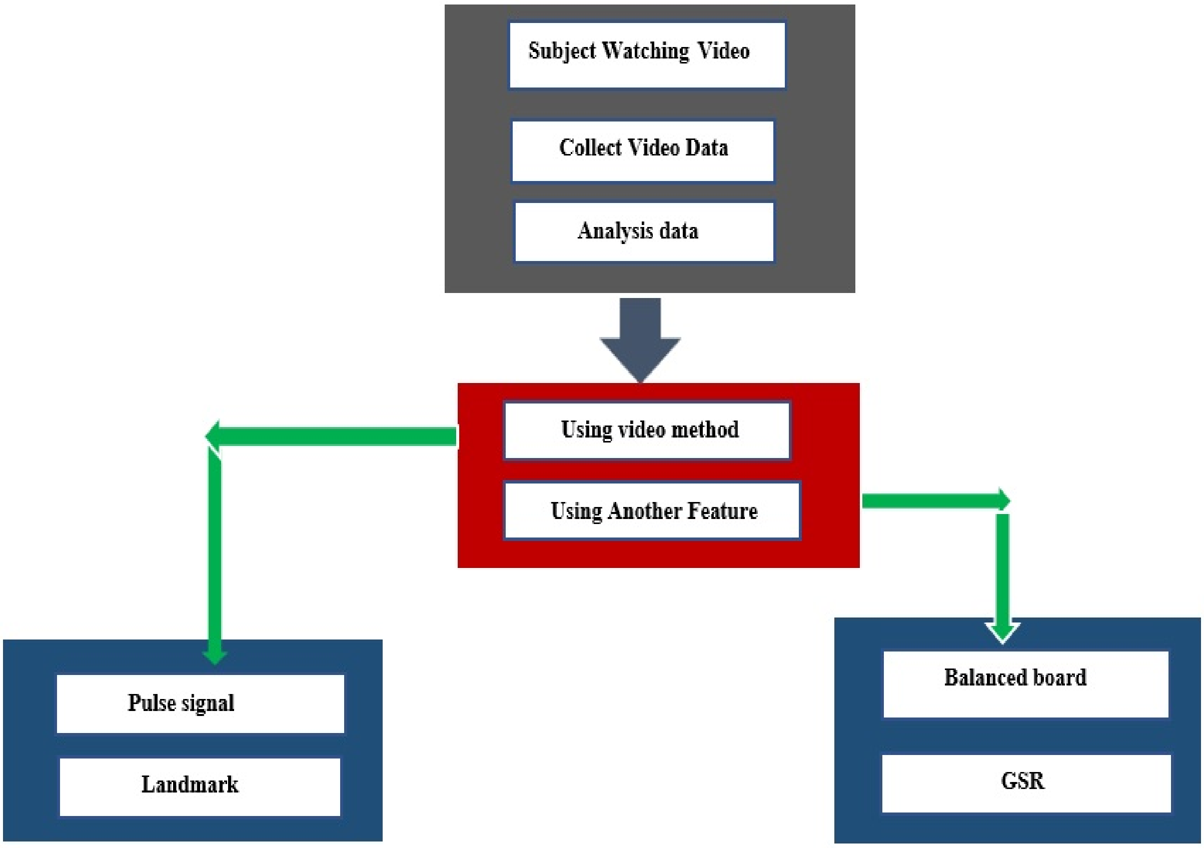


Figure 5.2: System Overview

human emotions has attracted a lot of interest in the recent years. Emotion evaluation is a problem, which is not easy to overcome. However, sometimes people are show their false emotions. Some people are always trying to hide their real emotions. For human emotion recognition various types of health monitoring systems are like a variable wearable sensor to develop portable devices that can continuously monitor and process several vital body parameters. In this part, we work on some wearable device for physical and emotional changes monitoring. The device obtains user's key physiological signals: video method, balanced board, pulse oximeter blood pressure and skin conductance and derives the user's emotion states as well.

5.3 Experiment Setup

For this work, we select undergraduate and postgraduate engineering students from the Saitama University were recruited for the data collection. They were all Asian descent and in sound health. The experiment consisted of three parts. The participants aged from 20 years to 25 years and there were 25 males and 15 females. they all of Asians and they were completely in normal and healthy state. Then using a Sony 4K FDR-AX30 camera, we recorded 29 FPS videos of each subject's face for 15 minutes while the subject watched a game video. After that, we recorded the subject for 10 minutes while he or she watched a boring video clip. In all cases, we determined the average HR for each minute of the viewing sessions. We make a video clip for 30 minutes. first 15 minutes excite game video clip, and then 5 minutes relax music video song & last 10 minutes boring clips for excited and boring. We record the face video for all subject for while the subject is watching the movie video clips. And in the same time all subject wear wearable sensor pulse oximeter, Wii fit balanced board and shimmer GSR. For One subject we have excited, relax & boring (15+5+10) minutes data.

Figure 5.3: Data collection flow chart

We calculated for each subject for the two conditions heart rate data compare their normal heart rate data. Total all subject 15 minutes excited heart rate data from HR video measurement method and wearable sensor data. And 10 minutes boring heart rate data from HR video measurement method and wearable sensor. Find out all subject normal value & high value for two condition (excited & boring). After collecting our videos, we chose 1 minutes of the videos where the peak emotional responses would be expected. For example, in the case of the comedy video clip, we chose the 30 second segment of the video with the punchline for the joke. In the case of the horror video clip, we chose the 30 second segment with the scariest part of the video. We then used the Open Face facial landmark tracker [19] to estimate facial landmarks for each subject on the 30 second segments in both the comedy and horror cases. We then used the tracked facial landmarks in conjunction with our version of Lam and Kuno's remote PPG algorithm for estimating cardiac pulses. We also used Open Face to obtain intensities for 17 facial action units [41] for each frame in the videos we analyzed. These facial action units were obtained to use as a means of comparison between our cardiac based approach and a facial expression-based approach. For each 1 minutes clip, we determined the average intensity of each facial action unit resulting in 17 values associated with each 1 minutes video segment to be used for emotion recognition.

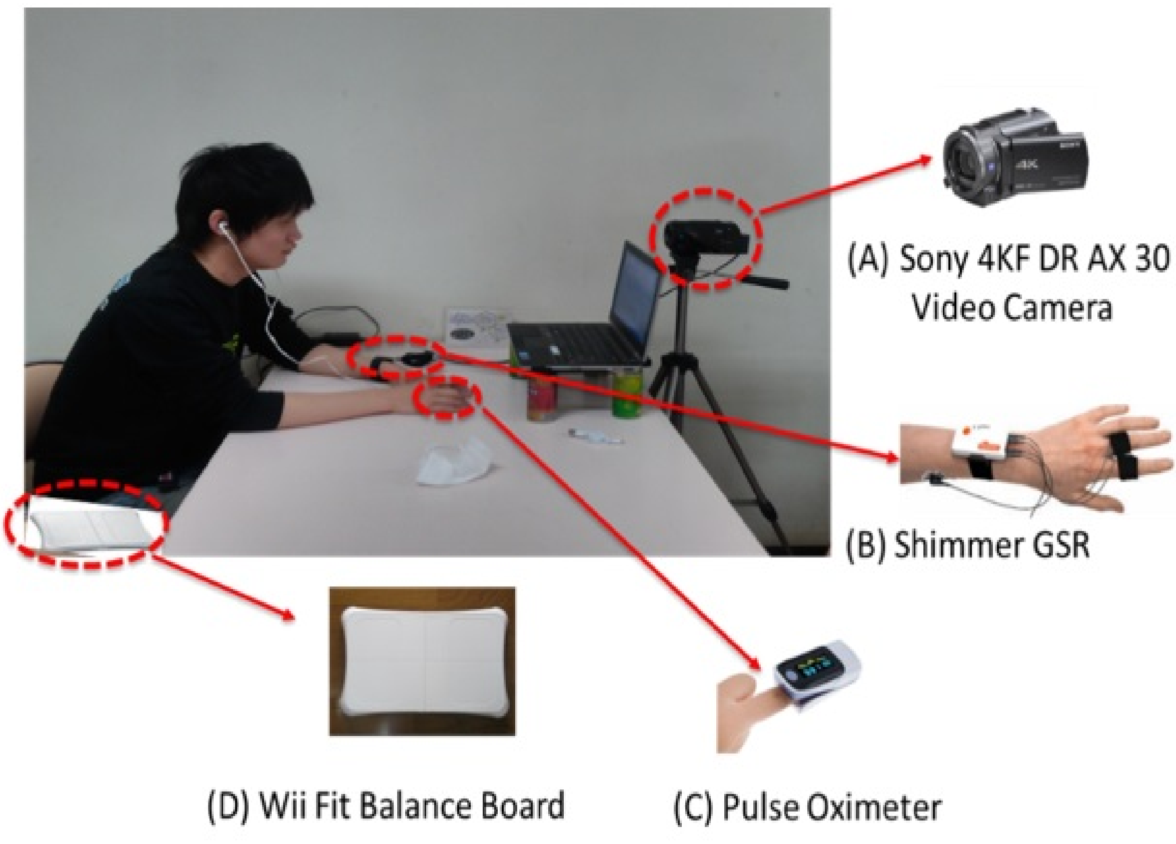


Figure 5.4: Subject watching video for experiment

Today we have many wearable devices, such as mobile phones and wearable sensors to measure physiological or behavioral data in our daily lives. This paper also aims to use technology to recognize stress levels using data from the devices that users always carry and wear. when one-person wear wearable sensor this person data automatically stores in devices and collect data from computer.in this part we use pulse oximeter, Wii fit balanced board and shimmer GSR as a wearable sensor. Pulse oximetry is a noninvasive method for monitoring a person's oxygen saturation (SO2). In its most common application mode, a sensor device is placed on a thin part of the patient's body, usually a fingertip or earlobe, or in the case of an infant, across a foot. The Wii Balance Board is shaped like a household body scale, with a plain white top and light gray bottom. Wii Balance Board, with four directional switches instead of pressure sensors. shift of center of pressure (CoP) is an indirect measure of postural sway and thus a measure of a person's ability to maintain balance. The GSR+ unit is suitable for measuring the electrical characteristics or conductance of your skin, as well as capturing an Optical Pulse/PPG (Photoplethysmogram) signal and converting to estimate heart rate (HR), using the Shimmer ear clip or optical pulse probe. The Galvanic Skin Response Sensor is used for real time GSR Biofeedback. The Shimmer GSR+ sensor monitors skin conductivity between two reusable electrodes attached to two fingers of one hand.

5.4 Experiment Results

Figure 5.5 show only one subject video screen shot when watching video clips. Total 30 minutes watching excited, relax and boring video clips. Thus, we have presented an approach to emotion recognition that is based on physiological responses rather than facial expressions. As a result, the emotions we detect cannot be easily faked. In addition, these cardiac pulse signals were entirely estimated from videos captured by a conventional RGB camera so no special equipment is required in our setup. Essentially, we have a system that operates completely using only computer vision techniques. A drawback of the current study is that our dataset does not have human subjects that intentionally tried to hide or fake their emotions and so it was not ideal for our tests. The results suggest that in situations where the human subjects have no reason to hide their emotions, facial expressions are still very reliable. In the first phase of our work, we showed video content to human subjects and collected HR data using an attached sensor pulse oximeter, Wii fit balanced board and skin conductance devices GSR for three emotional states (normal exited, relax, and boring) and the Figure 5.6 shows our video method HR data and wearable sensor (oximeter, balanced board ,GSR) physiological changes data.

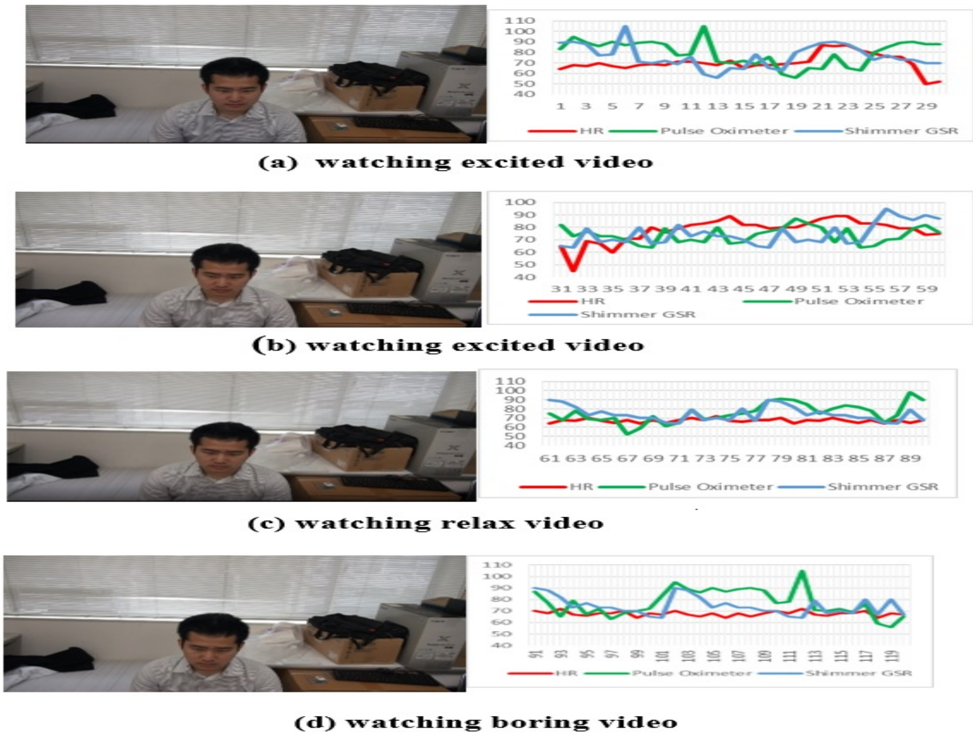


Figure 5.5: Different wearable sensor detect emotional changes

Remote photoplethysmography (PPG) algorithms has received attention more for emotion changes for human. These techniques allow one to read cardiac activity such as heart rates (HR) from conventional cameras by typically observing small changes in skin color over time. This thesis aims to use the sensed cardiac activity from remote PPG to detect and recognize emotions. Since remote PPG has been shown to work with conventional cameras, our proposed approach has the benefit that cameras such as web cams, surveillance cameras, and cellphone cameras could be used. With the ability to see cardiac activity without contact sensors, we present a convenient system for detecting inner emotions. Like in the psychophysiology literature, we chose to evaluate our approach by recognizing emotional reactions to emotionally stimulating videos such as horror and comedy clips. Figure 5.6 shows the analysis of Vision Based Video Method (VBVM) HR estimation and wearable sensors HR from video clips. We compare the HR data from VBVM and the wearable sensors. We found that the all methods or wearable sensor give us similar results from two emotional situations. Figure 5.6 also shows the mean HRs (over the entire viewing session for each subject) for the VBVM and wearable sensors are consistent across different subjects. Thus, the VBVM provides comparable performance to the wearable sensors. We collected

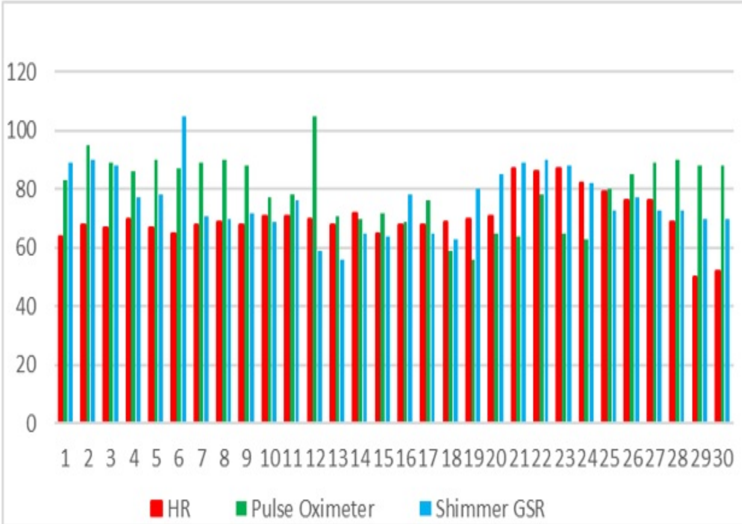


Figure 5.6: Compare our video HR data with wearable sensor

all subjects' HR data using VBVM and the wearable sensors. We observed them while they watched video clips and compared their resting HRs with their HRs while viewing the videos. Heart rate features play an important role in viewer interest detection. Common sources of the heart rate signal are chest and wrist (with photoplethysmogram (PPG)). Mean of heart rate and range of heart rate acceleration and deceleration have been found useful in measuring emotion and effort-related appraisals. Heart rate and skin response and pulse oximeter and balanced board found to different subjects' different types of changes. Figure 5.7 show only five subjects' changes rate for their data when they are watching 30 minutes video clips work on physiological signals measures emotions and affects in viewing content. Physiological signals such as heart rate, respiration, and skin response potentially reflect changes in underlying attitude or mood of viewers. Heart rate and heart rate variability (HRV) have been extensively used for emotion and affect assessment. Heart rate is inversely proportional to valence and heart rate increases with pleasantness and HRV decreases with excited, relax, and boring.

After watching video all of subjects we ask them fill a from which is emotional status related. Figure 5.8 show the graph of all subjects for their emotional changes rate when they are watching video clips. Excited video they feel 75%, relax video70% and boring video 65% they feel their feeling.it is little bit difficult to show one-person real feeling. videos content are all subjects are not same feeling they are feel. Just we try to combine their feeling and video clips.

Number	Modality	Feature	
1	HR	F1	F4,F1
	Wii Fit	F2	
	Oximeter	F3	
	GSR	F4	
2	HR	F1	F3,F4
	Wii Fit	F2	
	Oximeter	F3	
	GSR	F4	
3	HR	F1	F4,F2
	Wii Fit	F2	
	Oximeter	F3	
	GSR	F4	
4	HR	F1	F1,F3
	Wii Fit	F2	
	Oximeter	F3	
	GSR	F4	
5	HR	F1	F2,F1
	Wii Fit	F2	
	Oximeter	F3	
	GSR	F4	

Figure 5.7: Features extracted from wearable sensors and heart rate data

5.5 Chapter Summery

In this chapter, we compare our vision-based video method for heart rate and pulse signal data with various types of wearable sensors. Though our experiments we make a video clips for three types of emotional states. And as a mansion earliar watching video period we collect their emotional states.in our method HR which collect from video and another physiological changes data collect from werable sensor. Such information about emotional changes through viewer responses, which are either visually observable or physiological signals, both of which include physiological signs and heart rate take the sensor data while the participants are watching a movie of exiting, relaxed, and boring video. Experiments results confirm that our video-based method can be a useful dataset for further emotions recognition study.

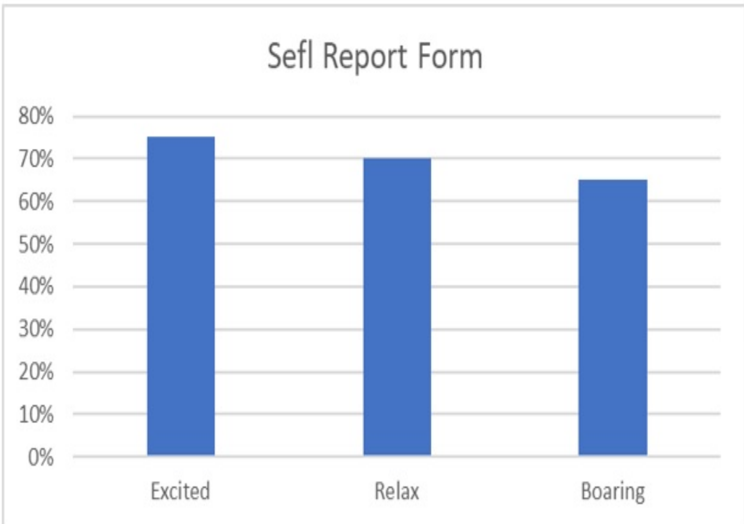


Figure 5.8: All subjects Self report after watching video

Chapter 6

Conclusions and Future Work

6.1 Conclusions

We have presented an approach to emotion recognition that is based on physiological responses rather than facial expressions. As a result, the emotions we detect cannot be easily faked. In addition, these cardiac pulse signals were entirely estimated from videos captured by a conventional RGB camera so no special equipment is required. Essentially, we have a system that operates completely using only computer vision techniques.

A drawback of the current study is that our dataset does not have human subjects that intentionally tried to hide or fake their emotions and so it was not ideal for our tests. However, we were able to show that even with a naive approach like taking the estimated cardiac pulse signals, performing dimensionality reduction using PCA, and then leaning via linear SVM, we achieved surprisingly good accuracy.

In this paper, we collected HR data for three emotional states (normal resting, funny, and horror) and performed a statistical analysis of the results. The HR data received using the VBVM is strongly correlated with the wearable sensor ground truth data. Moreover, our experimental results show that the HRs for the funny clips increase approximately 24.04% and there is a 19.06% rise for horror clips. We were able to detect emotional state changes from HR.

Another interesting line of future work is that we will develop an approach to automatically decide whether a person has experienced inner emotional change. The current work here establishes the ground work by statistically verifying the feasibility of the basic sensors used in our framework but we currently have no automated way to decide if someone is reacting to emotional stimuli. (For example, a Support Vector Machine or other machine learning technique could be used to take in various HR data and classify people's reactions.)

Moreover, we will account for additional emotional states like sadness, anger, and frustration. Differentiating between the various types of emotions will likely require some form of detailed subtle facial expression analysis but detecting changes in HR would still be needed to indicate the presence of emotional stimulation as a first step.

We also plan to continue our work to evaluating multiple people at the same time (as is done in the demonstration at the end of the paper). This would be particularly interesting as we could then observe the emotional reactions of audience members in movie screenings. This would be useful for pilot screenings so that movie producers could adjust content to improve the quality of movies. An interesting idea might even be to have a system that could adaptability alter the movie's content based on the sensed emotions in order to enhance the emotional experience of watching a movie.

6.2 Future Work

In the future, we will continue to investigate better ways to extract relevant features from the cardiac pulse signals to improve accuracy. Given the fact that our system only requires a conventional RGB camera, it would be interesting to explore various applications in effective computing in the future. Our method has some limitations like extreme changes in illumination and rapid motion that affect accuracy. In the future, we will resolve these problems to improve our system's performance. In addition, we will also investigate using the cardiac PPG signal itself and computing metrics such as HR variability, which is known to also be a good indicator of emotional change. It would be interesting to see if observing the continuous changes in the cardiac PPG signal itself could provide even more information about a person's emotional state.

Publication List

[1] Keya Das, Antony lam, Yoshinori Kobayashi, Yoshinori Kuno, Detecting Inner Emotions From Video Based Heart Rate sensing, Lecture Notes in Computer Science, Vol 10363,pp48 to57,Springer,2017.

[2] Keya Das, Antony lam, Yoshinori Kobayashi, Yoshinori Kuno, "Classification of Emotions from Video Based Cardiac Pulse Estimation", Lecture Notes in Computer Science, Springer, 2018.

[3] Keya Das, Kouyou Otsu, Antony lam, Yoshinori Kobayashi, Yoshinori Kuno, Towards Detecting the Inner Emotions of Multiple People,99th SICE,2017.

[4] Antony lam Kouyou Otsu, Keya Das, and Yoshinori Kuno, 'Towards Taking Pulses Over YouTube to Determine Interest in Video Content." In International Workshop on Frontiers ofComputer Vision (IW-FCV), IEEE, 2018. (Oral)

[5] Kouyou Otsu, Tomoki Kurahashi, Keya Das, Hisato Fukuda, Antony Lam, Yoshinori Kobayashi, Yoshinori Kuno, Heart rate measurement method using video images robust to environmental changes, The 23rd Symposium on Sensing via Image Information (SSII2017),2017(poster).

[6] Kouyou Otsu, Keya Das, Hisato Fukuda, Antony Lam, Yoshinori Kobayashi, Yoshinori Kuno, Robust and fast heart rate measurement method based on video analysis, The 24rd Symposium on Sensing via Image Information (SSII2018),2018.

Bibliography

[1] Ira Cohen, Nicu Sebe, Ashutosh Garg, Lawrence S. Chen, and Thomas S. Huang, “Facial Expression recognition from video sequences”, Multimedia and Expo, ICME ’02. Proceed-ings. 2002 IEEE International Conference,2002.

[2] Shiqing Zhang, Xiaoming Zhao, and Bicheng Lei, “Facial Expression recognition based on local binary patters and local fisher discriminant analysis”, PMC, 2011

[3] R. M. Stern, W. J. Ray and K. S. Quigley, “Psychophysiological Recording,” 2nd Edition, Oxford University Press, New York, 2001.

[4] Stern, J.A., 1964. Toward a definition of psychophysiology. Psychophysiology, 1(1), pp.90-91.

[5] K.H. Kim, S.W. Bang, S.R. Kim, ”Emotion recognition system using short-term monitoring of physiological signals,” Medical and Biological Engineering and Computing, vol. 42, pp. 419-427, 2004

[6] Cong Zong and Mohamed Chetouani, ”Hilbert-Huang transform based physiological signals analysis for emotion recognition,” IEEE International Symposium on Signal Processing and Information Technology, pp. 334-339, 2009.

[7] R.W. Picard, E. Vyzas, and J. Healey, ”Toward machine emotional intelligence: analysis of affective physiological state,” IEEE Transactions on Pattern Analysis and Machine Intelli-gence, vol. 23(10), pp. 1176-1189, June. 2001.

[8] Guillaume Chanel, Joep J.M. Kierkels, Mohammad Soleymani, and Thierry Pun, ”Short-term emotion assessment in a recall paradigm,” International Journal of Human-Computer Studies, vol. 67, pp. 607-627, 2009

[9] Xiaobai Li, Jie Chen, Guoying Zhao, and Matti Pietkainen, “Remote heart rate measurement from face videos under realistic situations”, CVPR ’14 Proceedings of the 2014 IEEE Con-ference on Computer Vision and Pattern Recognition, PP 4264-4271,2014

[10] Hamed Monkaresi, M Sazzad, and Rafael A Calvo, "Using remote heart rate measurement for Affect Detection", Florida Artificial Intelligence Research Society Conference, The Twenty-Seventh International Flairs Conference, PP 119-123, 2014

[11] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. T. Freeman, "Eulerian vid-eo magnification for revealing subtle changes in the world". ACM Trans. 2012.

[12] S. Kwon, H. Kim, and K. S. Park. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In IEEE Engineering in Medicine and Biology So-ciety (EMBC), pages 2174–2177, Aug 2012

[13] Verkruysse, W., Svaasand, L.O., Nelson, J.S.: Remote plethysmographic imaging using ambient light. Optics Express16(2008) 21434–21445

[14] Antony Lam and Yoshinori Kuno, "Robust Heart Rate Measurement from Video Using Se-lect Random Patches", pp 3640-3648., ICCV 2015

[15] M.-Z. Poh, D. McDuff, and R. Picard. Advancements in noncontact, multiparameter physio-logical measurements using a webcam. IEEE Transactions on Biomedical Engineer-ing,58(1):7–11, Jan 2011

[16] Xiaobao Li, Jie Chen, Guoying Zhao, Matti Pietikainen. Remote heart rate measurement from face videos under realistic situations. In IEEE Computer Vision and Pattern Recognition (CVPR). (2014) 4264–4271

[17] Sergey Tulyakov, Xavier Alameda-Pineda, Elisa Ricci, Lijun Yin, Jeffrey F. Cohn, Nicu Sebe. Self- adaptive matrix completion for heart rate estimation from face videos under realistic conditions. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2016)

[18] Kanitthika Kaewkannate and Soochan Kim, "A comparison of wearable fitness devices", Kaewkannate and Kim BMC Public Health (2016).

[19] Kelly R. Evenson, Michelle M. Goto and Robert D. Furberg. Systematic review of the validity and reliability of consumer-wearable activity trackers, International Journal of Behavioral Nutrition and Physical Activity,2015.

[20] Guha Balakrishnan, Fredo Durand, John Guttag, Detecting Pulse from Head Motions in Video, IEEE Computer Vision and Pattern Recognition (CVPR), 2013.

[21] W. Verkruysse, L.O. Svaasand, J.S Nelson, “Remote plethysmographic imaging using ambient light”, Optics Express 16, pp. 21434–21445,2008.

[22] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J.F. Cohn, N. Sebe, “Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR 2016.

[23] G.Balakrishnan, F.Durand, J.Guttag, “Detecting Pulse from Head Motions in Video”,Proceedings of the IEEE Computer Vision and Pattern Recognition,CVPR 2013.

[24] P.R.Chakraborty, L. Zhang, D. Tjondronegoro, and V. Chandra, “Using viewer’s facial expression and heart rate for sports video highlights detection”,pp. 371-378., ACM 2015

[25] B. Amos, B. Ludwiczuk, M. Satyanarayanan, “Openface: A general-purpose face recognition library with mobile applications”, CMU-CS-16-118, CMU School of Computer Science, Tech. Rep, 2016.

[26] P.Ekman and E. L. Rosenberg, “ What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)”,New York: Oxford University Press,1997.

[27] Antony Lam, Kouyou Otsu, Keya Das, and Yoshinori Kuno, “Towards Taking Pulses Over YouTube to Determine Interest in Video Content”, Proceedings of the IEEE International Conference on Computer Vision (IW-FCV), IEEE, 2018.

[28] G. de Haan and V. Jeanne, “Robust Pulse Rate from Chrominance-Based rPPG”, IEEE Transactions on Biomedical Engineering, vol. 60, no. 10, pp. 2878-2886, Oct. 2013.

[29] Keya Das, Sarwar Ali, Koyo Otsu, Hisato Fukuda, Antony Lam, Yoshinori Kobayashi, and Yoshinori Kuno, “Detecting Inner Emotions from Video Based Heart Rate Sensing”, 13th International conference on intelligent computing, pp. 48-57, ICIC 2017.

[30] Assel Davletcharova,Sherin Sugathan, “Detection and Analysis of Emotion From Speech Signals”, Procedia Computer Science , 2015

[31] Choubeila Maaoui and Alain Pruski, "Emotion Recognition through Physiological Signal for Human-Machine Communication", Vedran Kordic (Ed.), ISBN: 978-953-307-062-9, Cutting Edge Robotics 2010.

[32] Uriel Martinez-Hernandez, Imran Mahmood, A.Dehghani-Sanij, "Simultaneous Bayesian Recognition of Locomotion and Gait Phases with wearable sensors", Proceedings of the IEEE International Conference on Computer Vision (IW-FCV), IEEE, 2018

[33] M.Shamim Hossain and Ghulam Muhammad, "An Emotion Recognition System for Mobile Applications", Vol. 2169-3536, pp. 2281-2287, IEEE, 2017.

[34] Atefeh Goshvarpour, Ataollah Abbasi, Ateke Goshvarpour, "An accurate emotion recognition system using ECG and GSR signals and matching pursuit method", Biomedical Journal 40(355-368) 2017.

[35] Johnston DW, Anastasiades P, "The relation between heart rate and mood in real life", 1990;34(1):21-7. US National Library of Medicine, 2011.

[36] Khalili, Z. and Moradi, M.H. (2009), "Emotion Recognition System Using Brain and Peripheral Signals: Using Correlation Dimension to Improve the Results of EEG",IJCNN 2009, 1571-1575, 2009

[37] Raj Rakshit, V Ramu Reddy et al, " Emotion detection and recognition using HRV features derived from photoplethysmogram signals" , ERM4CT Article No.: 2 Pages 1–6, 2016.

[38] Revathi Priya, "Emotion Recognition from Physiological signals using Biosensors", Submitted for Research Seminar on Emotion Recognition, 2012.

[39] Yun-Kyung Lee ; Jun Jo ; Yongkwi Lee ; Hyun Soon Shin ; Oh-Wook Kwon, "Particle filter-based noise reduction of PPG signals for robust emotion recognition", ICCE,2012

[40] Ross A. Clark et al, "Validity and reliability of the Nintendo Wii Balance Board for assessment of standing balance", http://dx.doi.org/10.1016/j.gaitpost.2009.11.01, 2009.

[41] Keya Das, Antony Lam,Hisato Fukuda, Yoshinori Kobayashi, and Yoshinori Kuno," Classification of Emotions from Video Based Cardiac Pulse Estimation", ICIC 2018: pp 296-305, 2018.

[42] Nithya Roopa.S, “Emotion Recognition from Facial Expression using Deep Learning “, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6S, August 2019

[43] D Y Liliana, “Emotion recognition from facial expression using deep convolutional neural network”, International Conference of Computer and Informatics Engineering (IC2IE) IOP Conf. 2018, Series: Journal of Physics: Conf. Series 1193 (2019) 012004.

[44] Carlos Busso, Zhigang Deng el al, “Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information”,

[45] Black, M. J. and Yacoob, Y., “Tracking and recognizing rigid and non-rigid facial motions using local parametric model of image motion.”, In Proceedings of the International Conference on Computer Vision, pages 374–381. IEEE Computer Society, Cambridge, MA, 1995.

[46] Essa, Pentland, A. P.,” Coding, analysis, interpretation, and recognition of facial expressions”, IEEE Transc. On Pattern Analysis and Machine Intelligence,19(7):757–763, JULY 1997.

[47] Ekman, P., Friesen, W. V., “Facial Action Coding System: A Technique for Measurement of Facial Movement.” Consulting Psychologists Press Palo Alto, California, 1978.

[48] Mase K., “Recognition of facial expression from optical flow”. IEICE Transc., E. 74(10):3474–3483, 0ctober 1991.

[49] Pantic, M., Rothkrantz, L.J.M., “Toward an affect-sensitive multimodal human-computer interaction”., Proceedings of the IEEE, Volume: 91 Issue: 9, Sept. 2003. Page(s): 1370 – 1390

[50] Tian, Ying-li, Kanade, T. and Cohn, J., “Recognizing Lower Face Action Units for Facial Expression Analysis”., Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG & # 39;00), March, 2000, pp. 484 – 490.

[51] Yacoob, Y., Davis, L.,” Computing spatio-temporal representations of human faces”., Computer Vision and Pattern Recognition, 1994. Proceedings CVPR & # 39;94., 1994 IEEE Computer Society Conference on, 21-23 June 1994 Page(s): 70 –75.

[52] Stanley E. Jones,Curtis D. LeBaron, "Research on the Relationship Between Verbal and Nonverbal Communication: Emerging Integrations", 10 January 2006

[53] Yang, P., "Intercultural nonverbal communication competence as intercultural responsiveness in the second language learning classroom. In J. Kathryn & amp; R. M. Jason (Eds.), Intercultural responsiveness in the second language learning classroom", IGI Global, pp.127-147, 2017

[54] Alessandro Vinciarelli el al, "Towards a Technology of Nonverbal Communication: Vocal Behavior in Social and Affective Phenomena", IGI, pp. 133- 156.2010

[55] Priyanka Chettupuzhakkaran ; N Sindhu, "Emotion Recognition from physiological Signals Using Time-Frequency Analysis Methods", ICETIETR, 2018

[56] Gouizi K1, Bereksi Reguig F, Maaoui C., "Emotion recognition from physiological signals"., J Med Eng Technol. 2011 Aug-Oct;35(6-7):300-7

[57] Fernando Alonso-Martín, María Malfaz et al, "A Multimodal Emotion Detection System during Human-Robot Interaction", Sensors (Basel). 2013 Nov; 13(11): 15549–15581.

[58] Weihua Cao, Luefeng Chen, et al, "Facial Expression Emotion Recognition Based Human-robot Interaction System", Senior Member,IEEE.

[59] Human–robot interaction, Wikipedia

[60] Isabel Bush, "Measuring Heart Rate from Video", Stanford Computer Science 353 Serra Mall, Stanford, CA 94305

[61] Husam Salih, Lalit kulkarni ," Study of Video based Facial Expression and Emotions Recognition Methods" , International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), At Palladam, India, 2017, DOI: 10.1109/I- SMAC.2017.8058267

[62] Samad, Rosdiyana, and Hideyuki Sawada. ,& quot;Edge based Facial Feature Extraction Using Gabor Wavelet and Convolution Filters.& quot; In MVA, pp. 430-433. 2011.

[63] Thai, Le Hoang, Nguyen Do Thai Nguyen, and Tran Son Hai., & quot;A facial expression classification system integrating canny, principal component analysis and artificial neural network.& quot;, International Journal of Machine Learning and Computing, Vol. 1, No. 4, 2011, 388-393,2011

[64] Sisodia, Priya, Akhilesh Verma, and Sachin Kansal., & quot;Human Facial Expression Recognition using Gabor Filter Bank with Minimum Number of Feature Vectors. & quot;, International Journal of Applied Information Systems, Volume 5 – No. 9, July 2013 pp. 9-13., 2013.

[65] Meher, Sukanya Sagarika, and Pallavi Maben., & quot;Face recognition and facial expression identification using PCA.", In Advance Computing Conference, 2014 IEEE International, pp. 1093- 1098. IEEE, 2014.

[66] Jun Yu & amp; Zengfu Wang, "A Video-Based Facial Motion Tracking and Expression Recognition System", Springer Science Business Media New York 2016.

[67] Aniruddha Dey, "Contour based Procedure for Face Detection and Tracking from Video", 3rd Int'I Conf. on Recent Advances in Information Technology I RAIT-20161, 2016.

[68] Le Nguyen Bao, Dac-Nhuong Le, Le Van Chung and Gia Nhu Nguye"Performance Evaluation of Video-Based Face Recognition Approaches for Online Video Contextual Advertisement User-Oriented System", Information Systems Design and Intelligent Applications pp 287-295 ,2016.

[69] Huang, Chien-Ming; Cakmak, Maya; Mutlu, Bilge (2015). Adaptive Coordination Strategies for Human-Robot Handovers (PDF). Robotics: Science and Systems.

[70] "WeBuild: Automatically Distributing Assembly Tasks Among Collocated Workers to Improve Coordination & quot; (PDF). 2017.

[71] Hentout, Abdelfetah; Aouache, Mustapha; Maoudj, Abderraouf; Akli, Isma (2019- 08-18). & quot;Human–robot interaction in industrial collaborative robotics: a literature review of the decade 2008–2017 & quot;. Advanced Robotics. 33 (15–16): 764–799. doi:10.1080/01691864.2019.1636714. ISSN 0169-1864.

[72] Aggogeri, Francesco; Mikolajczyk, Tadeusz; O'Kane, James (April 2019). & quot; Robotics for rehabilitation of hand movement in stroke survivors & quot;. Advances in Mechanical Engineering. 11 (4): 168781401984192. doi:10.1177/1687814019841921. ISSN 1687-8140.

[73] Ona, Edwin Daniel; Garcia-Haro, Juan Miguel; Jardon, Alberto; Balaguer, Carlos (2019-06-26). & quot;Robotics in Health Care: Perspectives of Robot-Aided

Interventions in Clinical Practice for Rehabilitation of Upper Limbs & quot;. Applied Sciences. 9 (13): 2586. doi:10.3390/app9132586. ISSN 2076-3417.

[74] Marian S. Bartlett. Face Image Analysis by Unsupervised Learning, volume 612 of The Kluwer International Series on Engineering and Computer Science. Kluwer Academic Publishers, Boston, 2001.

[75] M. Pantic and J.M. Rothcrantz. Automatic analysis of facial expressions: State of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12):1424–1445, 2000.

[76] Crane, E. A., Shami, N. S., and Peter, C. 2007. Let's get emotional: emotion research in human computer interaction. In ACM CHI '07 Extended Abstracts on Human Factors in Computing Systems. San Jose, CA, USA, April 28 - May 03. pp.2101-2104

[77] Herbon A., Oehme A., and Zentsch E., Jan 2006. Emotions in ambient intelligence–an experiment on how to measure affective states. zmms.tu-berlin.de.

[78] Money A.G., Agius H., 2008. Are Affective Video Summaries Feasible? Emotion in HCI: Joint Proceedings of the 2005, 2006, and 2007 International Workshops, page 142-149.

[79] Money A.G., Agius H., 2008. Video Playing with Our Emotions. Emotion in HCI: Joint Proceedings of the 2005, 2006, and 2007 International Workshops, page 168-171.

[80] Peter C., Beale R. (eds.): Affect and Emotion in Human Computer Interaction. LNCS, vol. 4868. Springer, Heidelberg (2008) (to appear).

[81] Emotion-in-HCI website (2008). http://www.emotion-inhci.net.