Title	人の顔の可視画像と顔温度画像解析による感情推定に関する研究
Author(s)	Nguyen, Viet Hung
Citation	
Issue Date	2015-06
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/12879
Rights	
Description	Supervisor:小谷　一孔，情報科学研究科，博士

Estimation of Human Emotion by Analyzing of Visible Expressions and Thermal Facial Images

by

Nguyen Viet Hung

submitted to
Japan Advanced Institute of Science and Technology
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Supervisor: Associate Professor Kazunori Kotani

School of Information Science
Japan Advanced Institute of Science and Technology

June, 2015

Abstract

In our day-to-day life, communication plays a very important role. Emotion is a convenient way for human to communicate. As a result, research on human emotion estimation has become a key focus area of Human Computer Interaction (HCI), Human Robot Interaction (HRI) and Computer Vision. There are lots of works done on this topic, and many promising approaches have been proposed. The current dominant approaches to human emotion estimation reply on visual-based signals that are on or over the skin. Through analyzing expression, they want to predict the emotions behind the expression. However, there exist several hard problems had not been solved well for real system to handle naturally occurring emotions such as uncontrolled environment, poker-face, fake emotion, deliberately displayed and exaggerated expression of prototypical emotions. Our work presents a novel framework for human emotion estimation based on fusion of visible-based and thermal-based signals to fill these gaps.

The motivation behind this effort is to capitalize on the permanency of innate characteristics that are under the face skin using thermal Infrared (IR) signals. To establish feasibility, we propose special methodologies, temperature spaces and thermal Regions of Interest (t-ROIs) for feature-based level, thermal Principal Component Analysis (t-PCA) and norm Eigen-space Method based on Class-features (n-EMC) for decision-based level, and fusion models of visible-based and thermal-based features. To conduct experiments, a multimodal facial emotion database (KTFE) with strict procedures is built. The positive experimental results show that the proposal framework has merit, especially with respect to the problem of poker-face and/or uncontrolled environment. More importantly, the results demonstrate the feasibility of fusion of visible-based and thermal-based in human emotion estimation and open the way to solve challenges for complex emotions.

Keywords: Estimation of Human Emotion, t-ROIs, t-PCA, n-EMC, Thermal Infrared Image, KTFE Database.

Acknowledgments

It has been a long and challenging journey from the shape of the first idea until the forming of this dissertation. Besides many tears of disappointment and ambiguity when getting lost in research, there was also a lot of happiness for achieving the desired results, for challenging myself, and for being able to keep hoping. I would not have overcome the obstacles in my road without the support of so many people around me, who helped me stand up after failures and made me firm with the road I have chosen.

First and foremost, I offer my sincerest gratitude to my supervisor, Associate Professor Kazunori Kotani, who has supported me throughout my study with his knowledge, guidance, and encouragement. Without his consistent support, I could not finish the work in this dissertation.

I want to thank my dissertation committee members, Professor Jianwu Dang, Professor Bao Ho, Associate Professor Hirokazu Tanaka, Associate Professor Bac Le for their valuable suggestions for my dissertation.

I am also grateful to Associate Professor Atsuo Yoshitaka, Associate Professor Nguyen Le Minh, Assistant Professor Fan Chen and all lab mates for their helpful discussion, encouragement and invaluable comments.

I also would like to give my special thanks to Professor Akira Shimazu and Professor Mizuhito Ogawa for their great efforts to build up FIVE-JAIST program. I would like to acknowledge the 322 Scholarship from Vietnam Ministry of Education and Training, FIVE-JAIST program, JAIST research grant, for supporting me during my doctoral study.

I would also like to acknowledge University of Pedagogy, HCM city (HUP), University of Science, Vietnam National University, HCM city (HUS), and Japan Advanced Institute of Science and Technology (JAIST), who gave me chance to join the FIVE-JAIST program and helped me to complete the program. My colleagues in HUP and HUS have provided me a good environment during my studying. Their support and encouragement helped me much to complete this dissertation.

Dedication

To my parents and youngest brother: Dr.Nguyen Viet Dong, Dao Thi Thu Ha and Nguyen Viet Thinh. Without their love, support and encouragement, I can not achieve anything. To my wife: Nguyen Hoang To Loan. Without Loan's support, I can not focus on finishing my dissertation.

Publications

[1] H. Nguyen, F. Chen, K. Kotani, B. Le: " Fusion of Visible Images and Thermal Image Sequences for Automated Facial Emotion Estimation, " Journal of Mobile Multimedia, vol. 10, no. 3 & 4, pp.294-308. (Nov.2014).

[2] H. Nguyen, F. Chen, K. Kotani, B. Le: "Human Emotion Estimation using Wavelet Transform and t-ROIs for Fusion of Visible Images and Thermal Image Sequences," The 14th International Conference on Computational Science and Its Applications, Guimarães, Portugal. (June. 2014) (**Best Paper Award**).

[3] H. Nguyen, K. Kotani, F. Chen,B. Le: "Estimation of Human Emotions Using Thermal Facial Information," Proc. SPIE 9069, Fifth International Conference on Graphic and Image Processing. (Jan. 2014).

[4] H. Nguyen, K. Kotani, F. Chen, B. Le: "A Thermal Facial Emotion Database and Its Analysis," The 6th Pacific-Rim Symposium, PSIVT 2013, Guanajuato, Mexico. (Oct, 2013).

[5] H. Nguyen, K. Kotani, F. Chen,B. Le: "Estimation of Human Emotion Using Regions of Interest in Facial Thermal Image," Proc.PCSJ/IMPS 2013. (Nov.2013).

[6] H. Nguyen: "Facial Feature Extraction Based on Wavelet Transform," Artificial Intelligence and Computational Intelligence.LNCS, vol. 5855, pp.330-339. (Nov. 2009).

[7] T.D.X. Duong, H. Nguyen: "Some Extension of Sparse Principal Component Analysis," International Journal of Machine Learning and Computing, Vol.2(6), no.5, pp.701-705. (Oct. 2012).

[8] L. Nguyen, Y. Kohda, K. Umemoto, H. Nguyen: "Value Creation in Auditing Service: A Framework for Analysis," The Third Asia Conference on Information Systems, 2014. (Dec.2014).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation of Research

There are many scientific research has proven that human beings use a variety of visual and auditory cues such as tone of voice, hand gestures, and face expression to express feeling, give feedback and understand other's emotion [1], [2], [3]. It seems easy, simple, casual and effortless for human to use these auditory and visual cues [4]. In spice of growing processing power and multiplicity of input-output modalities, to recognize, understand and interpret emotion are very difficult tasks for computer and robot [1]. In spice of that, there are many recent scholarly works on automated recognition, interpretation, and expression of emotion [1], [2], [5], [6]. Researchers still have inspiration to design and implement intelligent systems due to the potential of computers. Currently, many potential applications of intelligent computers and robots have been reported in some field such as Human-Computer Interaction (HCI), Robotics, and so on [7], [8], [9], [6], [10], [3], [11]. However those systems still have many limitation of recognizing, interpreter and expression emotions [7], [12], [13], [14],[17]. Therefore, research in this topic is still a big challenge.

In visual cues, one of the important source to understand emotion is face including facial expression [15]. Human face conveys to us such a wealth of social signals, and we are so expert at reading these. A further reason for interest in the face has been its biological background. The structure of the face has evolved to allow it to contain organs serving a range of functions; the mouth for eating, nose for breathing, eyes for seeing, ears for hearing,especially the signalling of emotion by movement of facial muscles [16].

A large number of researches relies on visual facial cues to recognize and classify the facial expressions and estimate emotion [96]. Not only we have academic researches but also hardware accessory, tools for implement visual-based Automatic Facial Expression Analysis (FEA) which are being developed, test and made available. Recently, some researches and applications have reach over 80 % accuracy in analyzing facial expression [96]. There are many surveys, discussion about strengths as well as limitation when using visual-based for FEA in underlying theories and implementation details[96], [35], [97], [95], [55].

Based on their discussion, most of researches of FEA are tested under controlled environments, laboratory environments. There is a lack of accurate and robust FEA methods to be deployed in uncontrolled environments. Ambient light intensity is one of the factors to reduce the effective of those researches.

Besides, inconsistency between facial expressions and human motions is an impossible mission with visual-based systems, for example, another might be very happy but his face does not express any changes.

Those limitation of using visual-based for FEA and human emotion estimation has inspired researchers to explore the other sources to fill out the gaps. Using non-visual signal is one of solutions. Recent works in variety fields such as computational intelligence, psychology, physiology, neuropsychology, pattern recognition, machine learning and HCI show a useful, effective in using of non-visual signal to solve the problem of FEA, Human emotion estimation [32], [111], [78], [17].

There are many human bio-physiological signals, which are considered useful for estimating human emotion, are used to conduct on human emotion estimation, FEA. With bio-physiological signal, to obtain them, systems need to directly contact with human body, therefore, systems using non-visual cues remain intrusive . Many discussion and reviews of using non-visual cues for FEA, human emotion estimation in theory and implementation are showed in [17]. Because of intrusive manner, which is major operational difference with visual-based system, the FEA and human emotion estimation using non-visual based are still having some obstacles. Even though there exist some problems of using non-visual based cues, recent advances in thermal infrared (IR) image made it possible to acquire a very useful human bio-physiology signals, body temperature, through non-intrusive and non-contact means [18]. Human skin can be measure through IR image in non-invasive, non-contact, and illumination manner [19], [20], [17]. It is also considered as a function of thermal-muscular, hmodynamics and metabolic factors [21], [17]. There are many studies using the facial hmodynamics variations and thermal features to detect transition of emotional states [22], to classify affects and facial expressions [23],[24],[25],[26],[27],[28],[29], [30], [31], [32].

Most of researches used only thermal images or skin temperature inferring from pixel gray-level of thermal images [25], [33]. The database, which they used to experiment, includes just static thermal images. With emotions, sequence of thermal images seem to give the best knowledge to estimate human emotions. The emotion/expressions including those databases are deliberate, extraordinary expression,unreal emotion and overplay features, fake emotion. There are very few attentions using both of visible and thermal information to recognize facial expressions or human emotions.

Motivated by the success of previous studies and based on their disadvantages, in our work, firstly, we propose and establish a thermal facial expression and emotion database to allow the research in facial expression analysis to be more realistic; secondly, this work explores the possibilities of estimating human emotions using visible information and sequence of thermal information. Our works are done on the assumption that visible expression and thermal IR information are caused by the same emotion.

1.2 Objective of Research

Research in FEA and human emotion estimation has been biased toward the visible spectrum for a variety of reasons. Among those is the availability and low cost of visible cameras and the undeniable fact that face expression/emotion is one of the activities of the human visual system [86]. Although there are a significant number of sophisticated algorithms and advanced computational methods for implementing the visible-based FEA and human emotion estimation system [96], [97],[55], [35], FEA and especially human emotion estimation using thermal-based cues have received relatively little attention compared to those systems[111], [32].

This work focuses on developing an human emotion estimation system using visible-based cues and thermal-based cues.

The work began by investigating, designing, proposing, and building an multi-modal spontaneous visible and thermal facial thermal database. This database will help researches in this field to have more realistic database to make experiment. Some analysis based on database shows the relation between visible expressions and true emotion.

To reduce the disadvantages of thermal-based information, some methods are proposed. And to understand the effective of fusion method of visible-based and thermal-based information, an fusion method is proposed.

1.3 Contributions and Achievements

One innovation in this study has been using thermal IR information to estimate human emotion and integrating features extracted from thermal IR image and visible imaging to assess peoples emotional state. The main contributions of this research have been:

i) To produce new multi-modal, spontaneous visible and thermal facial emotion database which helps researches in this field to have more realistic database.

ii) To fill out the gaps of visible and thermal IR image, some techniques are proposed such as, removing the effect of eyeglass, thermal Regions of Interest(t-ROIs).

iii) To propose a fusion feature-level fusion and decision-level fusion from visual-based and thermal-based signals to estimate human emotion. This is one of the first attempts in human emotion estimation field using both visual-based and thermal-based signals.

1.4 Thesis Layout

The layout of this thesis is arranged in the following manner:

This document comprises of 7 chapters and a list of references.

Chapter 2 first discusses the existing and potential approaches using visible-based and non-visible based signal. Strengths and limitations of existing FEA and human emotion estimation systems are also examined.

Chapter 3 We analyze the procedure to obtain the database, propose and establish the multi-model spontaneous visible and thermal facial emotion database. We also present some analysis of that database.

Chapter 4 We propose a method to fill out the gap of thermal IR information. With our method, the negative effect of eyeglasses is reduced.

Chapter 5 We propose thermal Regions of Interest (t-ROIs) for thermal IR information. Using t-ROIs, we can reduce the effect of eyeglasses, increase the accuracy rate and decrease the running time.

Chapter 6 We propose a fusion method using both visible-based and thermal-based information. A fusion method includes feature-level fusion and decision-level fusion.

Chapter 7 provides a summary of this work and future works.

Reference section is appended at the end of this document. It provides a list of cited work.

Chapter 2

Review of Human Emotion Estimation Using Visible-Based and Non-Visible Based Signals

The human face has a fascinating capacity to express emotions. It conveys to us such a wealth of social signal, and we are so expert at reading these [16]. However, this is a big challenge for computer to understand emotion as well as we do. In order to make human-computer interaction effective, many studies have been investigating the possibilities of developing affective HCI models [3], [17]. Therefore, FEA and human emotion estimation have emerged as important research areas during the last three decades [34].

A survey of existing FEA and human emotion estimation based on visible-based and non-visible based signal is presented in the following paragraphs.

2.1 Review of Automatic Human Emotion Estimation Using Visual-Based Signals

In our definition, human emotion estimation system using visual-based signals is facial expression analysis system,because emotion signals obtained from visible cues are expressions. Face expression and emotion estimation have been a focus of research in human behavior for a hundred years [36]. Mehrabian [37] found that when people are communicating feelings and attitudes, 55% of the message is conveyed through facial expression alone, vocal cues provide 38% and the remaining 7% is via verbal cues. Other studies confirmed that facial visual signals are used during Human-Human Communication [14], help understand cognition and behavior [38], and are believed to have a significant role in the HCI and HRI systems [39], [40].

There are many applications of facial expression from variety of fields, for example, marketing [41], perceptual user interface, Human-Robot Interaction [42], [43],[44], drowsy driver detection [45], tele-nursing [46], pain assessment [47], analyzing mother-infant interaction [48], autism [49], social robotic [50], , facial animation [52], [53] and expression mapping for video gaming [54], [55] There tables, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, show some prominent researches in this fields.

To estimate the human emotions, most of approaches try to map the draw data to other spaces which have better describe, and are easy to estimate or classify the human emotions. Feature extraction or representation is a good maps by finding the proper features for the best representation of human emotions. And to estimate human emotion based on these features, classification is another stage to divide features into several categories.

2.1.1 Feature Extraction and Representation

Feature extraction and representation are the key to extract the useful information of face. We can categorize facial feature extraction into two main categories: appearance-feature based and geometric-feature based methods. Geometric features are the shape and locations of facial components such as eyes, eyebrows, nose, cheek-bones, mouth. From the obtained facial components or facial feature points, a feature vector which are

Reference	Feature Extraction	Classifier	Database	Sample size	Performance	Important Points
Tian et al., [57]	Permanent features: Optical Flow, Gabor Wavelets and Multi-state Models. Transient features: Canny edge detection	2 ANNs, one for upper face and one for lower face	Cohn-Kanade and Ekman-Hager Facial Action Exemplars	Upper Face: 50 sample sequences from 14 subjects performing 7 AUs Lower Face: 63 sample sequences from 32 subjects performing 11 AUs	Recognition of Upper Face AUs: 96.4% Recognition of Lower Face AUs: 96.7% Average of 93.3% when generalized to independent databases	Recognizes posed expressions. Real-time system. Automatic Face detection. Head motion is handled Invariant to scaling. Uses facial feature tracker to reduce processing time.
Bourel et al. [58]	Local spatio-temporal vectors obtained from the EKLT tracker	Modular classifier with data fusion. Local classifiers are rank weighted kNN classifiers. Combination is a sum scheme.	Cohn-Kanade	30 subjects 25 sequences for 4 expressions (total of 100 video sequences)	Refer fig. 2.1	Deals with recognizing facial expressions in the presence of occlusions. Proposes use of modular classifiers instead of monolithic classifiers. Classification is done locally and then the classifier outputs are fused.
Pardas and Bonafonte, [59]	MPEG-4 FAPs extracted using an improved Active Contour algorithm and motion estimation	HMM	Cohn-Kanade	Used the whole DB	Overall efficiency of 84% (across 6 prototypic expressions) Experiments with: joy, surprise and anger: 98%, joy, surprise and sadness: 95%	Automatic extraction of MPEG-4 FAPs Proves that FAPs convey the necessary information that is required to extract the emotions.

Table 2.1: A summary of some of the posed and spontaneous expression recognition systems [35].

Reference	Feature Extraction	Classifier	Database	Sample size	Performance	Important Points
Cohen et al. [60]	A vector of extracted Motion Units (MUs) using PBVD tracker	NB, TAN, SSS, HMM, ML-HMM	Cohn-Kanade and own DB	Cohn-Kanade: 53 subjects Own DB: 5 subjects	Refer table [2.7]	Real-time system. Emotion classification from video. Suggests use of HMMs to automatically segment a video into different expression segments.
Cohen et al. [61]	A vector of extracted Motion Units (MUs) using PBVD tracker	NB, TAN and SSS	Cohn-Kanade and Chen-Huang	Refer table [2.9]	Refer table [2.8]	Real-time system. Uses semi-supervised learning to work with some labeled data and large amount of unlabeled data.
Bartlett et al. [62]	Gabor Wavelets	SVM, AdaSVM (SVM with AdaBoost).	Cohn-Kanade	313 sequences from 90 subjects. First and last frame used as training images	SVM with Linear Kernel: Automatic face detection: 84.8%; Manual alignment: 85.3%. SVM with RBF Kernel: Automatic face detection: 87.5%; Manual alignment: 87.6%	Fully automatic system. Real-time recognition at high level of accuracy Successfully deployed on Sony's Aibo pet robot, ATR's RoboVie and CU Animator
Michel and Kaliouby, [63]	Vector of feature displacements (Euclidean distance between neutral and peak)	SVM	Cohn-Kanade	For each basic emotion, 10 examples were used for training and 15 examples were used for testing.	With RBF Kernel: 87.9%. Person independent: 71.8% Person dependent (train and test data supplied by expert): 87.5% Person dependent (train and test data supplied by 6 users during ad-hoc interaction): 60.7%	Real-time system. Does not require any preprocessing.

Table 2.2: (continued) A summary of some of the posed and spontaneous expression recognition systems [35].

Reference	Feature Extraction	Classifier	Database	Sample size	Performance	Important Points
Pantic and Rothkrantz, [64]	Frontal and Profile facial points	Rule based classifier	MMI	25 subjects	86% accuracy	Recognizes facial expressions in frontal and profile views. Proposed a way to do automatic AU coding in profile images. Not Real-time.
Buciu and Pitas, [65]	Image representation using Non negative Matrix Factorization (NMF) and Local Non negative Matrix factorization (LNMF).	Nearest neighbor classifier using CSM and MCC	Cohn-Kanade and JAFFE	Cohn-Kanade: 164 samples JAFFE: 150 samples	Cohn-Kanade: LNMF with MCC gave the highest accuracy of 81.4% JAFFE: Only 55% to 68% for all 3 methods	PCA was also performed for comparison purpose. LNMF outperformed both PCA and NMF whereas NMF produces the poorest performance. CSM classifier is more reliable than MCC and gives better recognition.
Pantic and Patras, [66]	Tracking a set of 20 facial fiducial points	Temporal Rules	Cohn-Kanade and MMI	Cohn-Kanade: 90 images MMI: 45 images	Overall an average recognition of 90%	Recognizes 27 AUs. Invariant to occlusions like glasses and facial hair. Shown to give a better performance than the AFA system
Zheng et al., [67]	34 landmark points converted into a Labeled Graph (LG) using Gabor wavelet transform. Then a semantic expression vector built for each training face. KCCA used to learn the correlation between LG vector and semantic vector.	The correlation that is learnt is used to estimate semantic expression vector which is then used for classification	JAFFE and Ekman's Pictures of Affect	JAFFE: 183 images Ekman's: 96 images Neutral expressions were not chosen from either database	Using Semantic Info: On JAFFE DB: with Leave one image out (LOIO) cross validation: 85.79%, with Leave one subject out (LOSO) cross validation: 74.32%, On Ekman's DB: 81.25% Using Class Label Info: On JAFFE DB: with LOIO: 98.36%, with LOSO: 77.05%, On Ekman's DB: 78.13%	Used KCCA to recognize facial expressions The singularity problem of the Gram matrix has been tackled using an improved KCCA algorithm.

Table 2.3: *(continued)* A summary of some of the posed and spontaneous expression recognition systems [35].

Reference	Feature Extraction	Classifier	Database	Sample size	Performance	Important Points
Anderson and McOw-en, [68]	Motion signatures obtained by tracking using spatial ratio template tracker and performing optical flow on the face using multi-channel gradient model (MCGM)	SVM and MLP	CMU-Pittsburg AU coded DB and a non-expressive DB	CMU: 253 samples of 6 basic expres-sions. But these had to be preprocessed by reducing frame rate and scale Non-expressive: 10 subjects, 4800 frames long	Motion averaging using: co-articulation regions: 63.64%, 7x7 blocks: 77.92%, ratio template algorithms, with MLP: 81.82%, with SVM: 80.52%	Fully automated, multistage system. Real-time system. Able to operate efficiently in cluttered scenes. Used motion-averaging to con-dense the data that is fed to the classifier.
Aleksic and Katsaggelos, [69]	MPEG-4 FAPs, outer lip (group 8) and eyebrow (group 4) followed by PCA to reduce dimen-sionality	HMM and MS-HMM	Cohn-Kanade	284 recordings of 90 subjects	Using HMM: Only eye-brow FAPs: 58.8%, Only outer lip FAPs: 87.32%, Joint FAPs: 88.73% After assigning stream weights and then using a MS-HMM: 93.66% with outer lip having more weight than eyebrows.	Showed that performance im-provement is possible by using MS-HMMs and proposed a way to assign stream weights. Used PCA to reduce the dimen-sionality of the features before giving it to the HMM.
Pantic and Patras, [70]	Mid level parame-ters generated by tracking 15 facial points using particle filtering	Rule based classifier	MMI	1500 samples of both static and profile views (single and multiple AU activations)	86.6% on 96 test profile sequences	Automatic segmentation of input video into facial expres-sions. Recognition of temporal seg-ments of 27 AUs occurring alone or in combination. Automatic recognition of AUs from profile images.

Table 2.4: *(continued)* A summary of some of the posed and spontaneous expression recognition systems [35].

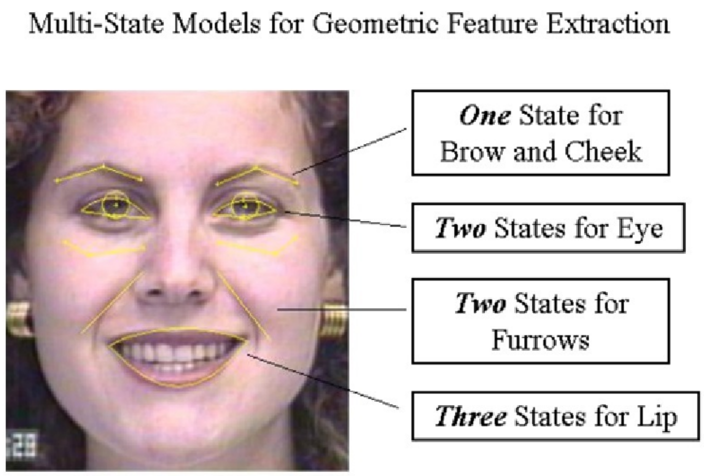


Figure 2.1: Geometric feature extraction [95]

formed, represents the face geometry. The appearance features include the appearance changes of human face.

■ Geometric feature

In general, geometric features include the facial shape, such as points, edges, curves, and their relative relations. The facial shape can be extracted by active shape models (ASMs), statistical models of shape of objects which performs energy minimization to deform a given shape to match the nearest salient contour [123]. The other approach, the facial shape can be defined by 58 facial landmarks or set of facial characteristic points around the eyes, eyebrows, mouth, chin and nose [62]. The common problem of these methods is mis-alignment problems, which is that localization of landmark can not be accurate enough. Fig 2.1 shows an example of geometric feature extraction.

■ Appearance feature

The appearance features present the skin texture changes of human face, such as wrinkles, furrows, eye area, mouth and nose. To extract the texture features, there are several popular methods, Gabor wavelets [124], and local binary pattern (LBP) [125].

• Gabor wavelets

A Gabor filter, named after Dennis Gabor, is a linear filter to detect the edges. In spatial domain, a family of Gabor kernel [122] is the product of a Gaussian envelope function and the relative width σ.

$$\Psi_{\rho,\gamma}(z) = \frac{| k_{\rho,\gamma} |^2}{\sigma^2} e^{-\|k_{\rho,\gamma}\|^2 / \sigma^2} \mid e^{ik_{\rho,\gamma} z} - e^{-\sigma^2/2} \mid \tag{2.1}$$

where $k_{\rho,\gamma}$ is frequency vector $k_{\mu,\gamma} = \frac{k_{max}}{g^{\gamma}} e^{i\Phi_\mu}$, $k_{max} = \Pi / 2, g = \sqrt{2}$, and $\Phi_\mu = \frac{\mu\Pi}{8}$, $z = (x, y)$. μ and γ are orientation and scale factors, respectively.

For each input image $I(x, y)$, the Gabor transformation at a particular position is computed as follows $G_{\rho,\gamma} = I(x, y) \times \Psi_{\rho,\gamma}(x, y)$.

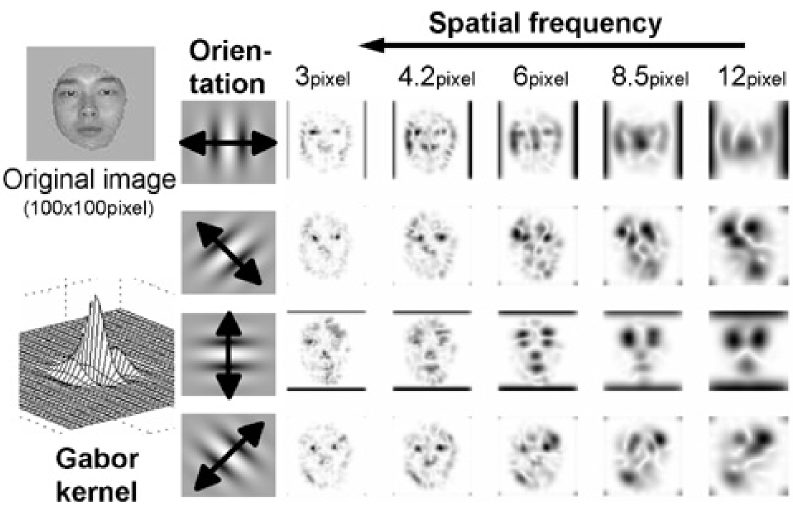


Figure 2.2: Gabor features extracted from face image [121]

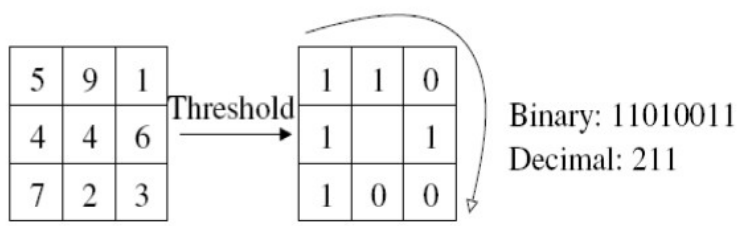


Figure 2.3: The LBP operator [125]

The feature vectors $F_{m,K}$ are given by

$$F_{m,n} = \sum_{i=x_n-m}^{x_n+m} \sum_{j=y_n-m}^{x_n+m} \mid G_{i,j} \mid \tag{2.2}$$

where $m = 2k+1, k \in N, n \in \overline{0,K}, \mid G \mid = \sqrt{Re\,(G)^2 + Im\,(G)^2}$ Fig2.2 shows Gabor features extracted from face image. • Local binary pattern features

The LBP are proposed by Ojala et al. [125]. The operator labels the pixels of an image by thresholding a 3×3 neighborhood of each pixel with the center value and considering the results as binary number 0 or 1 and m-bin histogram of the LBP labels computed over a region is used as a texture descriptor. The Fig.2.3 shows how the basic operator works. And LBP feature vectors extracted from each zone shows in Fig 2.4.

2.1.2 Classification

After extracting facial features, the classification/estimation methods are used in the last stage of an automatic facial expression analysis/emotion estimation system.

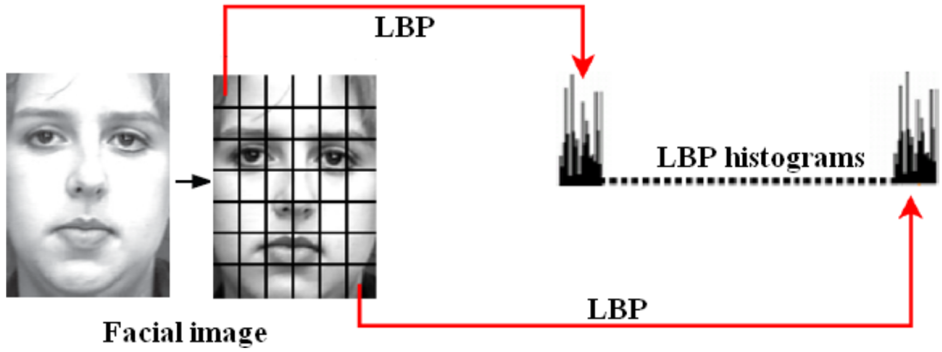


Figure 2.4: LBP feature vectors extracted from different region on the face image. [125]

• Linear Discriminant Analysis (LDA)

Linear discriminant analysis is one of the best know pattern classification by searching the projection axes on which different-classes' data points are far from each other while satisfying the same-class's data points to be close to each other [127].

Let F be a set of classes to be analyzed. Here, F is a set of all emotion classes. Assume that M_m^f in R^n training data are given as the facial pattern for each class $f \in F$ where $F = \{anger, disgust, fear, happiness, neutral, sadness, surprise\}$. Let Γ_i^f be the $i-th$ facial pattern where $i = \overline{1, M_f}$;

LDA finds a transformation matrix W that maps M_f points to N_n^f points in R^l $(l \leq \mid F \mid)$. The objective function of LDA is as follows:

$$max_w = \frac{w^\tau S_B W}{w^\tau S_W W} \tag{2.3}$$

where

$$M = \sum_{f \in F} M_f \tag{2.4}$$

$$\Psi^f = \frac{1}{M^f} \sum_{i=1}^{M_f} \Gamma_i^f ; \Psi = \frac{1}{M} \sum_{f \in F} \sum_{i=1}^{M_f} \Gamma_i^f \tag{2.5}$$

$$S_B = \frac{1}{M} \sum_{f \in F} M_f \|\Psi_f - \Psi\| \|\Psi_f - \Psi\|^\tau . \tag{2.6}$$

$$S_W = \frac{1}{M} \sum_{f \in F} \sum_{i=1}^{M_f} \|\Gamma_i^f - \Psi_f\| \|\Gamma_i^f - \Psi_f\|^\tau . \tag{2.7}$$

• Support Vector Machines (SVM)

Support vector machine is a supervised learning methods that construct a hyperplane or set of hyperplanes optimally separate the data points to one of two categories [126].

Let $Y_{train} = \{Y_1, Y_2, ..., Y_m\}$ is a set of m training data with target values given by $z_{train} = \{z_1, z_2, ..., z_m\}$, where $Y_i \in R^D$ and $z_i \in \{-1, 1\}; i \in \overline{1, m}$, the main task in training SVMs is to solve the following quadratic optimization problem:

Reference	Feature Extraction	Classifier	Database	Sample size	Performance	Important Points
Sebe et al., [71]	MUs generated from the PBVD tracker	Bayesian nets, SVMs and Decision Trees. Used voting algorithms like bagging and Boosting to improve results.	Created spontaneous emotions database. Also used Cohn-Kanade	Created DB: 28 subjects showing mostly neutral, joy, surprise and delight. Cohn-Kanade: 53 subjects	Using many different classifiers: Cohn-Kanade: 72.46% to 93.06%, Created DB: 86.77% to 95.57%. Using kNN with k = 3, best result of 93.57%	Recognizes spontaneous expressions Created an authentic DB where subjects are showing their natural facial expressions Evaluated several Machine Learning algorithms
Kotsia and Pitas, [72]	Geometric displacement of Candide nodes	Multiclass SVM: For expression recognition: used six-class SVM, one for each expression. For AU recognition: used one-class SVMs, one for each of the 8 chosen AUs used.	Cohn-Kanade	Whole DB	99.7% for facial expression recognition 95.1% for facial expression recognition based on AU detection	Recognizes either the six basic facial expressions or a set of chosen AUs. Very high recognition rates have been shown
Wang and Yin, [73]	Topographic context (TC) expression descriptors	QDC, LDA, SVC and NB	Cohn-Kanade and MMI	Cohn-Kanade: 53 subjects, 4 images per subject for each expression. Total of 864 images MMI: 5 subjects, 6 images per subject for each expression. Total of 180 images	Person dependent tests: on MMI: with QDC: 92.78%, with LDA: 93.33%, with NB: 85.56%, on Cohn-Kanade: with QDC: 82.52%, with LDA: 87.27%, with NB: 93.29%. Person independent tests: on Cohn-Kanade: with QDC: 81.96%, with LDA: 82.68%, with NB: 76.12%, with SVC: 77.68%	Proposed a topographic modeling approach in which the gray scale image is treated as a 3D surface. Analyzed the robustness against the distortion of detected face region and the different intensities of facial expressions.

Table 2.5: *(continued)* A summary of some of the posed and spontaneous expression recognition systems [35].

$$min_{\alpha} f\left(\alpha\right) = \frac{1}{2}\alpha^{\tau} Q \alpha - e^{\tau}\alpha \tag{2.8}$$

subject to

$$0 \leq \alpha_i \leq C, i \in \overline{1, m},$$

$$y^{\tau}\alpha = 0$$

Reference	Feature Extraction	Classifier	Database	Sample size	Performance	Important Points
Dornaika and Davoine, [74]	Candide face model used to track features.	First head pose is determined using Online Appearance Models and then expressions are recognized using a stochastic approach	Created own data	Used several video sequences. Also created a challenge 1600 frame test video, where subjects were allowed to display any expression in any order for any duration	Results have been spread across different graphs and charts. Interested readers can refer [24] to view the same.	Proposes a framework for simultaneous face tracking and expression recognition. 2 AR models per expression gave better mouth tracking and in turn better performance. The video sequences contained posed expressions.
Kotsia et al., [75]	3 approaches: Gabor features, DNMF algorithm and by Geometric displacement vectors extracted using Candide tracker	Multiclass SVM and MLP	Cohn-Kanade and JAFFE		Using JAFFE: with Gabor: 88.1%, with DNMF: 85.2% Using Cohn-Kanade: with Gabor: 91.6%, with DNMF: 86.7%, with SVM: 91.4%	Developed a system to recognize expressions in-spite of occlusions. Discusses the effect of occlusion on the 6 prototypic facial expressions.

Table 2.6: *(continued)* A summary of some of the posed and spontaneous expression recognition systems [35].

	Cohn-Kanade DB	Own DB
Person Dependent Tests	Insufficient data to perform person dependent tests	NB (Gaussian): 79.36% NB (Cauchy): 80.05% TAN: 83.31% HMM: 78.49% ML-HMM: 82.46%
Person Independent Tests	NB (Gaussian): 67.03% NB (Cauchy): 68.14% TAN: 73.22% Insufficient data to conduct tests using HMM and ML-HMM	NB (Gaussian): 60.23% NB (Cauchy): 64.77% TAN: 66.53% HMM: 55.71% ML-HMM: 58.63%

Table 2.7: Cohen et al.'s system's performance [59].

	Cohn-Kanade DB	Chen-Huang DB
Labeled and Unlabeled data	NB classifier: 69.10% TAN classifier: 69.30% SSS classifier: 74.80%	NB classifier: 58.54% TAN classifier: 62.87% SSS classifier: 74.99%
Only Labeled data	NB classifier: 77.70% TAN classifier: 80.40% SSS classifier: 81.80%	NB classifier: 71.78% TAN classifier: 80.31% SSS classifier: 83.62%

Table 2.8: Cohen et al.'s system's performance [60].

	Cohn-Kanade DB
Labeled and Unlabeled data	200 labeled, 2980 unlabeled for training and 1000 for testing
Only Labeled data	53 subjects displaying 4 to 6 expressions. 8 frames per expression sequence
	Chen-Huang DB
Labeled and Unlabeled data	300 labeled, 11982 unlabeled for training and 3555 for testing
Only Labeled data	5 subjects displaying 6 expressions. 60 frames per expression sequence

Table 2.9: Cohen et al.'s system's sample size [60].

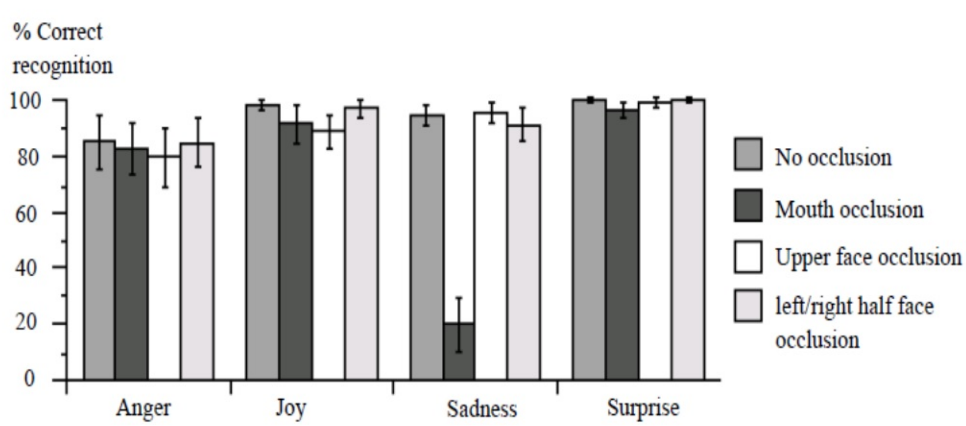


Figure 2.5: Bourel et al.'s systems performance. [57].

2.2 Review of Automatic Human Emotion Estimation Using Non-Visible Based Signals

Automatic human emotion estimation is an area with immense practical potential which includes a wide range of commercial and law enforcement applications, and it continues to be one of the most active research areas of computer vision. Even after over three decades of intense research, the state-of-the-art in human emotion estimation continues to improve, benefiting from advances in a range of different fields including image processing, pattern recognition, computer graphics and physiology. However, systems based on visible spectrum images continue to face challenges in the presence of illumination, pose and expression changes, as well as facial disguises, all of which can significantly decrease their accuracy. Among various approaches which have been proposed in an attempt to overcome these limitations, the use of infrared (IR) imaging has emerged as a particularly promising research direction[74].

The following paragraphs present a comprehensive and timely review of the literature on this subject.

2.2.1 Thermal Infrared (IR) Image

In Latin, 'infra' means "below", so "Infrared" means below red. The longest wavelengths of visible light is the red color. Therefore, Infrared wavelength is longer than the wavelength of red light visible.

The infrared ray is a form of an electromagnetic wave as well as a visual light or a radio wave. The wavelength band is 0.78 to 1000(μm) that is longer than visual light yet, shorter than radio wave, and the the wavelengths are classified from the near infrared to the far infrared region. However, it shall be considered as various classification have been proposed. The infrared ray is also energy radiated by motions of atoms and molecules on the surface of object, in case the temperature of object is more than absolute zero degree.

Estimation of facial emotions using different imaging modalities, specially thermal IR image is one of the hot research topics. There are some electromagnetic spectral bands such as X-rays and ultraviolet radiation which below the visible spectrum are harmful to the human body, hence cannot be employed for human face applications. However, thermal IR image is not harmful to the human body. Therefore it has been suggested as an alternative source of information for detection and recognition of faces and estimation of emotions.

The visible spectrum range of visual cameras are from 0.4 - 0.7 μm, while infrared spectrum range at 0.7 14.0 μm. The infrared spectrum comprises the reflected IR and the thermal IR wavebands. The IR spectrum is divided into four bands: the near-infrared (NIR) (0.7 - 0.9 μm) and the short-wave infrared (SWIR) (0.9 - 2.4 μm) spectra, the mid-wave infrared (MWIR) of the spectral range (3.0 - 5.0 μm) and long-wave infrared (LWIR) from (8.0 - 14.0 μm). This division of the IR spectrum is also observed in the manufacturing of IR cameras, which are often made with sensors that respond to electromagnetic radiation constrained to a particular sub-band. It should be emphasized that

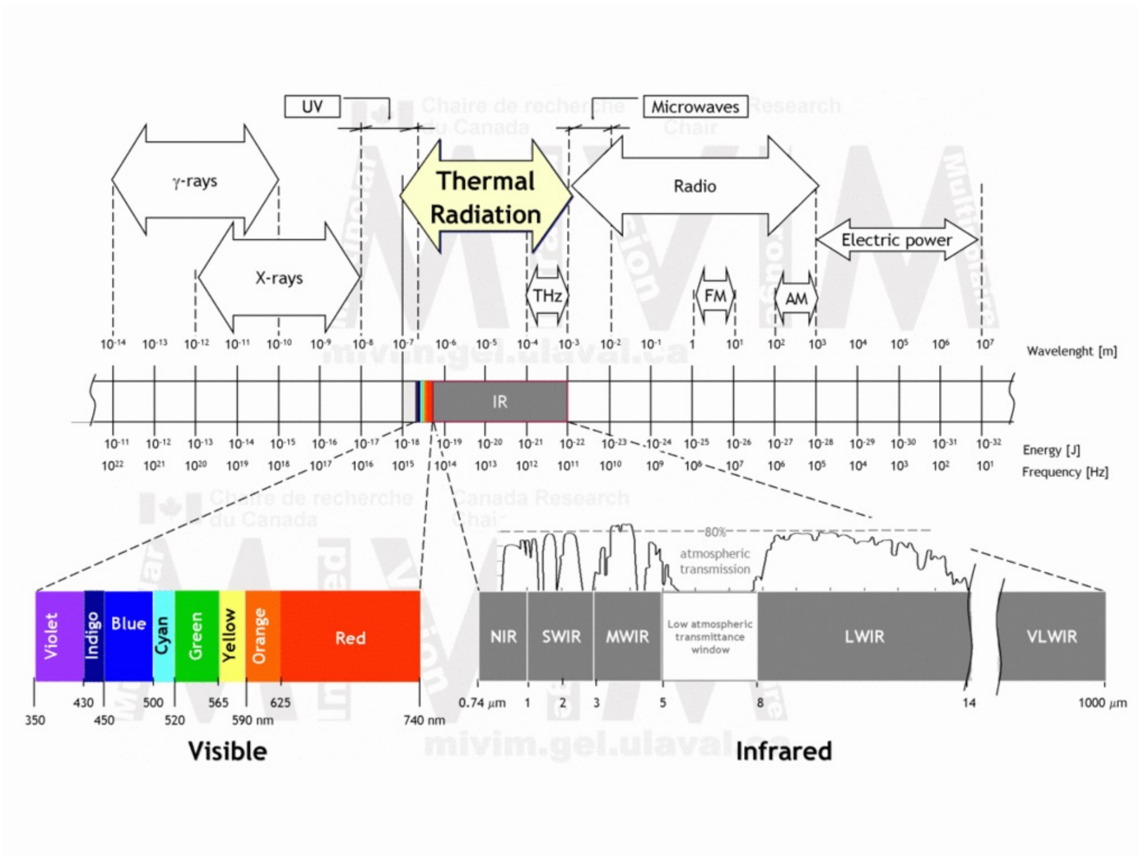


Figure 2.6: The infrared bands in the electromagnetic spectrum. Figure reprinted from [75]

the division of the IR spectrum is not arbitrary. Rather, different sub-bands correspond to continuous frequency chunks of the solar spectrum which are divided by absorption lines of different atmospheric gasses [76], [77].

In the context of human emotion estimation, one of the largest differences between different IR sub-bands emerges as a consequence of the human face and bodys heat emission spectrum. Specifically, most of the heat energy is emitted in LWIR sub-band, which is why it is often referred to as the thermal sub-band (this term is sometimes extended to include the MWIR sub-band). Significant heat is also emitted in the MWIR sub-band. Both of these sub-bands can be used to passively sense facial thermal emissions without an external source of light. This is one of the reasons why LWIR and MWIR sub-bands have received the most attention in the human emotion estimation literature. In contrast to them, facial heat emission in the SWIR and NIR sub-bands is very small and human emotion estimation systems operating on data acquired in these sub-bands require appropriate illuminators i.e. estimation is active in nature. And estimation of facial emotions in the thermal IR favors the LWIR due to much higher emissions in this band than in the MWIR. In recent years, the use of NIR also started received increasing attention from the human emotion estimation community, while the utility of the SWIR sub-band has yet [74].

2.2.2 Estimation of Human Emotion

Emotions of human can be obtained by expressions and skin temperature change. Whenever expressions change, facial skeletal muscles are active and produce heat for maintaining the body temperature [79], [80]. Using the facial EMG readings, scientists were able to discover an association between the muscular movements, muscle energy expenditure and the facial expressions of affective states [34], [85]. The major facial muscles that are considered responsive to emotions are shown in [2.3].

The increased blood volume flow under an area of facial skin (as the result of stress) is termed as reactive hyperemia [81]. Reactive hyperemia includes situations such as mechanical insult to the skin, chemical reactions causing vasodilatation of blood capillaries and thermal stress (like cold water immersion). Infrared imaging is used to diagnose, monitor and quantify hyperemia effects and quantify the dynamic stress on the skin [81], [82],[18]. Studies suggest that facial muscles either contract or expand when the facial expressions change [83]. Muscular contraction and expansion are believed to cause some fluctuations in the rate and volume of blood flow under the facial skin. A change in the emotional experience is also believed to influence the blood flow rate under the facial skin [19],[17],[84],[20].

There are some advantages when we use thermal IR images.

1. Facial skin temperature can be measured from a distance using the infrared cameras. Since no body contact is required, the target person may not notice any thermographic activity though this may result in breach of personal privacy and may raise some ethical issues. Despite these issues, surveillance and security communities require non-contact and secret monitoring of suspects and would benefit from thermal IR images based human emotion estimation system;

2. Modern infrared equipment allows non-invasive thermographic measurements. This may be particularly useful for medical and psychological diagnostic applications under

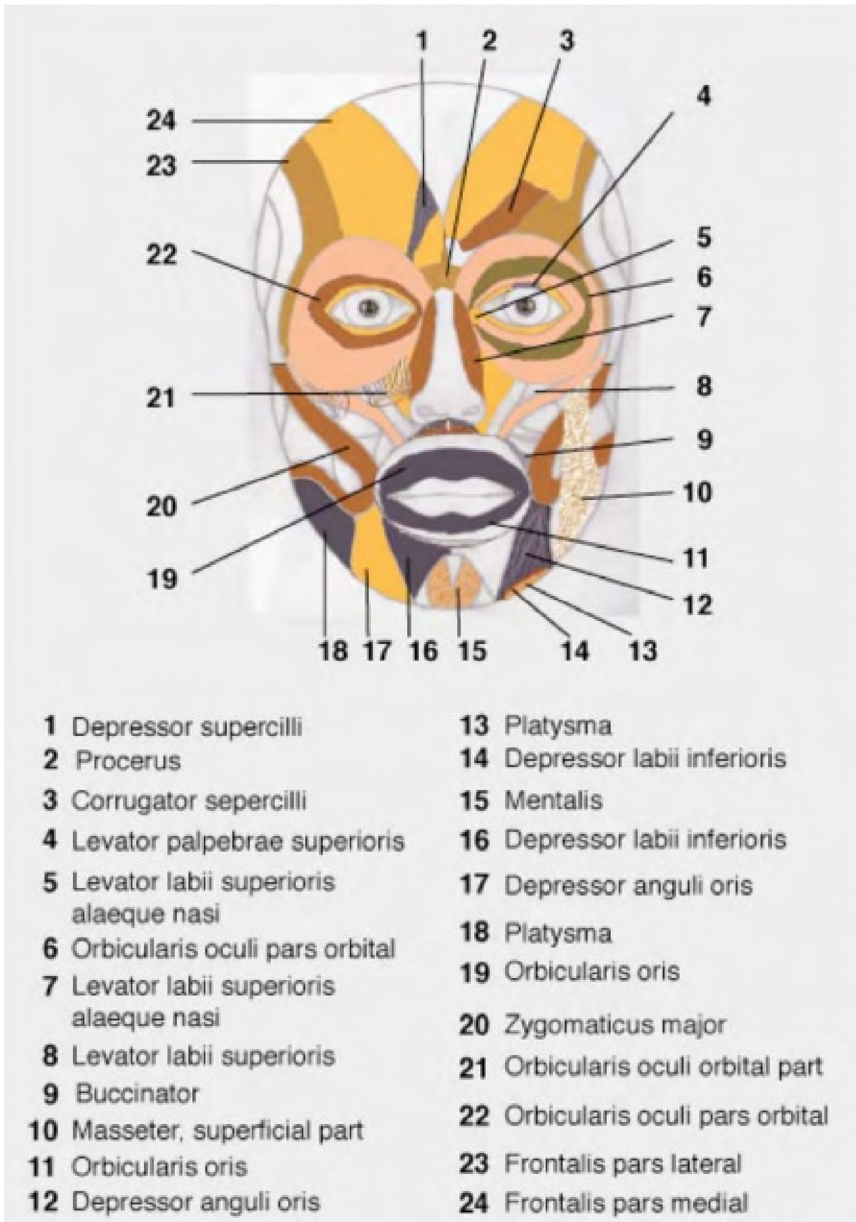


Figure 2.7: Frontal view of the facial muscle map showing all major facial muscles [17].

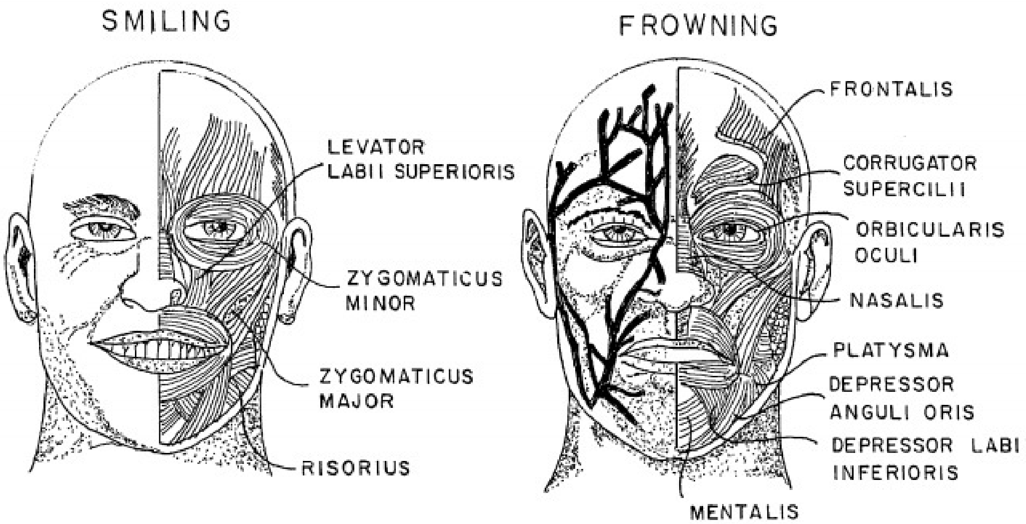


Figure 2.8: Images of vascular structure minutiae for an individual smiling and frowning, respectively [87].

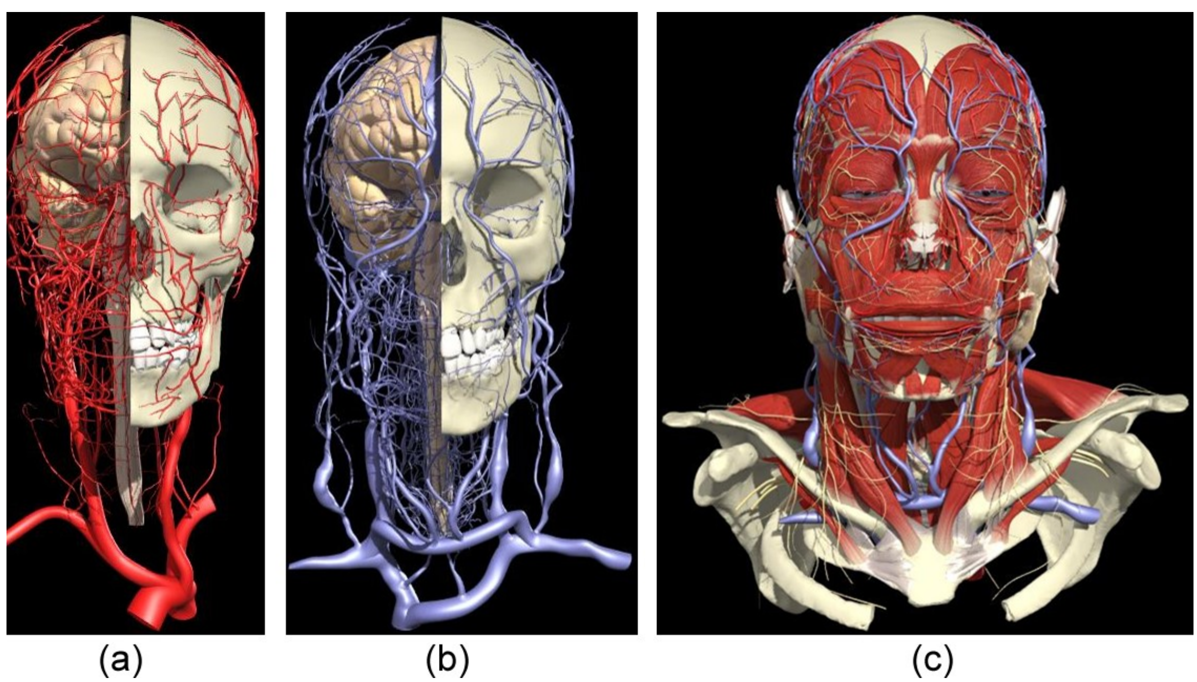


Figure 2.9: a) Overview of arterial network. (b) Overview of venous network. (c) Arteries and veins together under the facial surface [86].

conditions when patients are either unable or unwilling to cooperate;

3. Thermal IR images are invariant to light and illumination conditions;

4. Thermal IR image equipment is accessible and is becoming less expensive and affordable;

5. Modern thermal IR image equipment is light, aesthetically appealing and is easy to handle;

6. The latest infrared cameras are highly sensitive to any thermal variations on the human skin. These cameras are capable of sensing up to $\pm$ 0.05 ^{0}C thermal variations;

7. Thermal IR images provide both visual and physiological information for the human emotion estimation system;

8. Thermal IR images are safe and harmless to both the user of infrared equipment and the target individual [17].

Several approaches have been researched to classify human affective states using facial expressions and thermal imaging, a method to explore facial changes in temperature during the state assessment.

O.Kane et al. [88] have demonstrated noticeable changes in the thermal signature of the human face during breathing, muscle tension, aerobic exercise, and during aggressive playing of video games. These results suggest that thermal imaging is a promising technology that could be used to gain insight on the perception of human state assessment and also to understand underlying internal states.

Sophie Jarlier et al. [89] show that the thermal changes of the face are caused by the changes in the facial muscle contractions. The FACS coders are trained to produce different action unit combination at various intensities. These changes in action unit combination eventually cause the thermal patterns which can be classified using a PCA decomposition of the thermal signal and used K-nearest neighbor to classify seven expressions. The database for testing has four persons and the accuracy rate is 56.4 %. One of the things to be noted is that all the coders are forced to certain emotions; which makes it difficult to detect and characterize spontaneous expressions.

M.M.Khan et al. [23] suggested using Facial Thermal Feature Points (FTFPs), which are defined as facial points that undergo significant thermal changes in presenting an expression, and used Linear Discriminant Analysis (LDA) to classify intentional facial expressions based on Thermal Intensity Values (TIVs) recorded at the Facial Thermal Feature Points (FTFPs). The database has sixteen persons with five expressions and the accuracy rate ranges from 66.3% to 83.8%.

L.Trujillo et al. [24] proposed using a local and global automatic feature localization procedure to perform facial expression in thermal images. They used PCA to reduce the dimension and interest point clustering to estimate facial feature localization and Support Vector Machine (SVM) to classify three expressions.

B.Hernandez et al. [25] used SVM to classify the expressions surprise, happy, neutral from two inputs. The first input consists of selections of a set of suitable regions where the feature extraction is performed, second input is the Gray Level Co-occurrence Matrix used to compute region descriptors of the thermal IR images.

B.R.Nhan et al. [26] extracted time, frequency and time-frequency features from thermal infrared data to classify the natural responses in terms of subject-indicated levels of arousal and valence stimulated by the International Affective Picture System.

Y.Yoshitomi et al. [33] used two dimensional detection of temperature distribution on the face using infrared rays. Based on studies in the field of psychology, several blocks on the face are chosen for measuring the local temperature difference. With Back Propagation Neutral Network, the facial expression is recognized. The recognition accuracy reaches 90% with neutral, happy, surprising and sad expressions. However, the testing database is obtained from only one female frontal view. Y. Yoshimomi generated feature vectors by using a two-dimensional Discrete Cosine Transformation (2D-DCT) to transform the grayscale values of each block in the facial area of an image into their frequency components, and used them to recognize five expressions, including angry, happy, neutral, sad, and surprise. The mean expression accuracy is 80% with four test subjects [90].

Y.Koda et al. used the idea from [90] and added a proposed method for efficiently updating of training data, by only updating the training data with happy and neutral facial expression after an interval [91]. The expression accuracy increased from 80% to 87% with this new approach.

Wang et al. [32] proposed both decision-level and feature-level fusion methods using visible and thermal IR image. In feature-level, they used tools for the Active Appearance Model (AAM) to extract features and extracted three features of head motion for visible feature and calculated several statistical parameters including mean, standard deviation, minimum and maximum as thermal IR features. To select the feature, they used F-test statistic. They also used Bayesians networks (BNs) and SVMs to obtain the feature fusion. In decision-level, BNs and SVMs are used to classify three emotions, happiness, fear and disgust. The results show that their methods improved about 1.35% accuracy compare with only using visible features.

Yoshitomi et al. [22] proposed decision-level fusion of voices, visual and thermal IR image to recognize the affective states. DCT is used to extract the visible and thermal IR features, then two neutral networks are trained for obtained visible and thermal IR features, respectively. For voice recognition, Hidden Markov Models (HMMs) are used. To decide the results, simple weighted voting is used.

I. Pavlidis et al. [92] have shown evidence of a unique way to capture high definition thermal images of the face for detecting deceit. Exploring thermal images and detecting deceit has accuracy comparable to the polygraph examination. To attach electrodes and to perform a security screening at the airport using a polygraph mechanism for each individual person is almost impossible because of the amount of time needed. Using thermal imaging of a face gives a specific thermal signature for different emotions. In the paper[40], an experiment is conducted with twenty participants and they were asked to stab a mannequin, rob it for 20 dollar, and then prove that they are innocent. The thermal imaging was successful in correctly classifying 6 out of 8 participants who were guilty. Using thermal imagery as a stand-off sensor in turn helps to measure and analyze psychological responses without contact sensors.
Another study [93] measures the startling effect using thermal imaging. Facial thermal signatures changes have been seen near the periorbital and cheek regions for subjects after fright eliciting experiments. The study in [41] shows that thermal signatures of the face help us to determine the psychological state of a person. However, the above mentioned studies did not provide any pattern recognition analysis of thermal signatures to classify the emotions.

Liu and Wang [27] analyzed a facial temperature sequence data and computed statistical features and temperature difference histogram features. Further, Hidden Markov Models (HMM) were used to discriminate happiness, disgust and fear with a recognition rate of 68.11 %, 57.14% and 52.30%; respectively. The results also demonstrated the temperature information of the forehead is more useful than other regions of the face. They used samples from the USTC-NVIE (natural visible and infrared facial expression) database to evaluate their results [44]. All the research demonstrate that, thermal cameras could be used as a non-contact, non-invasive way to detect the changes in the temperature across the face [94].

Chapter 3

A Thermal Facial Emotion Database and Its Analysis

In recent years, thermal IR image has been extensively used in many fields such as military (e.g., target acquisition, surveillance, night vision, homing and tracking) and civilian purposes (e.g., medical diagnosis, thermal efficiency analysis, environmental monitoring). It is a promising alternative for investigation of facial expression and emotion. Currently there are very few database to support the research in facial expression and emotion, however most of them either only include posed thermal expression images or lack thermal IR information. For these reasons, we propose and establish a natural visible and thermal facial emotion database. The database contains spontaneous expression of 30 subjects. We also analyze a visible database, a thermal IR database to recognize expression and thermal IR information to recognize emotion.

3.1 Introduction

To choose the database which is used for testing the new system is one of the most important aspect of the developing any new detection, recognition, analysis and estimation system. If there is a common database used by all the researcher, then testing the new system, comparing it with the other state of art systems becomes a very easy and straightforward job. Therefore, creating or building a 'common' database or affective database which can satisfy some strict requirement of the problem domain and become a standard is a very difficult and challenging task.

With respect to emotion estimation or expression analysis or recognition, there are many novel databases which are created contained emotional expression and emotions described and surveyed in detail in [95],[96],[97]. However the problem of a standardized database for human emotion is still an open problem.

Currently most of database uses visible images or videos. However, under the lack of illumination, either darkness or exceeding of source of light, the result of visible expression analysis is not good. On the other hand, thermal IR images are not sensitive to light conditions. Consequently, using thermal IR images helps us to complete the gaps of visible images. Besides, the skin temperature changes are useful to classify the emotions [23] and facial expression is a good emotion-related behavior [98]. We can infer emotions from skin temperature and expressions from several special emotions. Moreover, most of the current databases used for research are visible and posed. The expressions, from those databases, are usually obtained from unreal emotion and overplay features. In addition, there are a few thermal facial image databases but they are posed thermal expression images. Even though a database is built in posed and spontaneous expressions, it still made some mistakes such as when they designed data acquisition, they forgot about time lag phenomenon or expressions are elicited by asking participants to imitate sample expressions, exaggerated expressions. With these reasons, we propose and establish a thermal facial expression and emotion database to allow the research in facial expression analysis to be more realistic.

In this chapter, we describe in detail the materials and methods to design and collect the thermal facial emotion database - KTFE (Kotani Thermal Facial Emotion) database. To verify the effectiveness of our spontaneous database, we analyze the effective of eliciting videos and use PCA (Principal Component Analysis), EMC (Eigenspace Method based on Class features) and PCA-EMC to classify facial expressions of a visible and thermal

Table 3.1: A summary of some of the facial expression databases that have been used in the past few years [35].

Database	Sample Details	Expression Elicitation and Data Recording Methods	Available Descriptions	Additional Notes	Reference
Cohn-Kanade Database (also known as CMU-Pittsburg database) [104], [105]	• 500 image sequences from 100 subjects • Age: 18 to 30 years • Gender: 65% female • Ethnicity: 15% African - Americans and 3% of Asians and Latinos	This DB contains only posed expressions. "The subjects were instructed to perform a series of 23 facial displays that included single action units (e.g. AU 12, or lip corners pulled obliquely) and action unit combinations (e.g. AU 1+2, or inner and outer brows raised). Each begins from a neutral or nearly neutral face. For each, an experimenter described and modeled the target display. Six were based on descriptions of prototypic emotions (i.e., joy, surprise, anger, fear, disgust, and sadness). These six tasks and mouth opening in the absence of other action units were annotated by certified FACS coders."	"Annotation of FACS Action Units and emotion-specified expressions"	"Images taken using 2 cameras: one directly in front of the subject and the other positioned 30 degrees to the subject's right. But the DB contains only the images taken from the frontal camera."	Information presented here has been quoted from [105]
MMI Facial Expression Database [106],[107]	• 52 different subjects of both sexes • Gender: 48% female • Age: 19 to 62 years • Ethnicity: European, Asian, or South American • Background: Natural lighting and variable backgrounds (for some samples)	This DB contains posed and spontaneous expressions. "The subjects were asked to display 79 series of expressions that included either a single AU (e.g., AU2) or a combination of a minimal number of AUs (e.g., AU8 cannot be displayed without AU25) or a prototypic combination of AUs (such as in expressions of emotion). Also, a short neutral state is available at the beginning and at the end of each expression." For natural expressions: "Children interacted with a comedian. Adults watching emotion inducing videos"	"Action Units, metadata (data format, facial view, shown AU, shown emotion, gender, age), analysis of AU temporal activation patterns"	"The emotions were determined using an expert annotator". A highlight of this DB is that it contains both frontal and profile view images.	Information presented here has been quoted from [108]

facial image database. We have also used PCA, EMC and PCA-EMC to classify emotions of thermal facial emotion database leading to very attractive results. Using the obtained thermal IR data, we can induce some interesting information relating to the temperature trends when emotions change.

3.2 Review of Existent Natural and Infrared Databases

Innumerable natural databases for facial expression analysis have been built since many years, such as Cohn-Kanade (CK) database also known as CMU-Pittsburg AU coded database [99]. This fairly extensive database has been used widely by facial expression analysis community. However the database contains only visible posed expression. Therefore, it is suitable to use for comparison against former systems, but not to use for spontaneous expression analysis and recognition. There are also several standardized databases containing both posed and spontaneous expressions, such as UA-UIUC [69],

Table 3.2: *(continued)* A summary of some of the facial expression databases that have been used in the past few years [35].

Database	Sample Details	Expression Elicitation and Data Recording Methods	Available Descriptions	Additional Notes	Reference
Spontaneous Expressions Database [71]	• 28 subjects	This DB contains spontaneous expressions. The subjects were asked to watch emotion inducing videos in a custom built video kiosk. Their expressions were recorded using hidden cameras. Then, the subjects were informed about the recording and were asked for their consent. Out of 60, 28 gave consent.	The database is self labeled. After watching the videos, the subjects recorded the emotions that they felt.	The researchers found that it is very difficult to induce all the emotions in all of the subjects. Joy, surprise and disgust were the most easy whereas sadness and fear were the most difficult.	Information presented here has been quoted from [71]
The AR Face Database [109], [110]	• 126 people • Gender: 70 men and 56 women • Over 4000 color images are available	This DB contains only posed expressions. "No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. Each person participated in two sessions, separated by two weeks time. The same pictures were taken in both sessions."	None	This database has frontal-faces with different expressions, illumination conditions and occlusions (scarf and sunglasses).	Information presented here has been quoted from [109]
CMU Pose, Illumination, Expression (PIE) Database [111]	• 41,368 images of 68 people • 4 different expressions.	This DB contains only posed expressions.	None	This database provides facial images for 13 different poses 43 different illumination conditions.	Information presented here has been quoted from [111]
The Japanese Female Facial Expression (JAFFEE) Database [112], [113]	• 219 images of 7 facial expressions (6 basic facial expressions + 1 neutral) • 10 Japanese female models.	This DB contains only posed expressions. The photos have been taken under strict controlled conditions of similar lighting and with the hair tied away from the face.	"Each image has been rated on 6 emotion adjectives by 92 Japanese subjects"	All the expressions are multiple AU expressions.	Information presented here has been quoted from [112], [113]

Table 3.3: Current thermal IR facial database.

Ref	Size	Wave band	Education	Lightning	Exp Des
NIST Equinox [115]	600 subjects 1919 infrared images	8-12μm 3-5μm	Posed	Above, left and right	Smiling, frowning, surprise
IRIS [116]	30 subjects, 4228 pairs of thermal and visible images	7-14μm	Posed	Left, right, both lights, dark	Surprise, laughing, anger
USTC-UVIE [117]	215 subjects	8-14 μm	Posed and spontaneous	Left, right and front	Happy, angry, neutral, disgusted, fearful, sad, and surprised

Figure 3.1: Current thermal IR facial database.

Ref	Size	Wave band	Education	Lightning	Exp Des
NIST Equinox [115]	600 subjects 1919 infrared images	8-12μm 3-5μm	Posed	Above, left and right	Smiling, frowning, surprise
IRIS [116]	30 subjects, 4228 pairs of thermal and visible images	7-14μm	Posed	Left, right, both lights, dark	Surprise, laughing, anger
USTC-UVIE [117]	215 subjects	8-14 μm	Posed and spontaneous	Left, right and front	Happy, angry, neutral, disgusted, fearful, sad, and surprised

MMI [64], AAI [108] and so on, co. A comprehensive survey of these databases is given in [65], [35].

Following the work done bey Sebe at al. [69], they found out that it is very difficult to induce the fear and sadness expression. They also found that spontaneous facial expression could be misleading such as some subjects had a sad expression when they were actually feeling happy. This raised a question on how we deal with those problems with the visible source. Is there any information to help us to reveal the hidden emotion? Currently, thermal IR infrared source is one solution. Besides, compared to the number of existing visible databases, only very few thermal face databases are available in the literature. Furthermore, these databases only include some posed thermal IR data and one spontaneous thermal IR data. In this document, we listed and compared several databases of infrared facial expression, along with the information related to the name, the number of subjects, wave band of thermal IR camera, lighting, illumination and expression description as table 3.3. Firstly, NIST Equinox [109] has been used in many researches of thermal IR image, which is not available anymore. Secondly, IRIS Thermal/Visible Face Database [110] is very useful only for face recognition because posed expressions are elicited by asking subjects to perform a series of emotional expressions in front of a camera. Thirdly, USTC-NVIE database [111] is a very good database and adaptable for a good posed and spontaneous thermal IR database. However, their procedure for data acquisition to induce emotions has a mistake. In their video clips to evoke emotion, the gaps between each emotion clip are 1-2 min long which is too short for participants to establish a neutral emotion status. They do not mention about the recording time before ending of each emotion clip. The changing of human temperate is later than the changing of emotion. Therefore, the time before ending of each emotion clips is very important.

In a short, there is only one facial expression database using visible and thermal IR image, although many expression databases use visible or thermal only. Furthermore, there exist several unclear and non-suitable procedures in these databases. These reasons motivated us to propose and build up another natural visible and infrared facial emotion

Table 3.4: Information of participants in building the KTFE database.

Number	Age	Sex	Education	Glasses	Nationality
2	32	2M	PostDoc	2 No	Viet
1	31	1M	PostDoc	1 No	Viet
2	30	1M, 1F	Phd	1 Yes, 1 No	Viet
1	29	1M	Master	1 Yes	Viet
6	28	3M, 3F	3Master, 3Phd	5 Yes, 1 No	Viet, Thai
1	27	1M	Master	1 No	Viet
5	26	3M, 2F	4Master, 1Phd	3 Yes, 2 No	Viet
5	25	4M, 1F	2Master, 3Phd	2 Yes, 3 No	Viet
6	24	4F,2M	Bachelor	4 Yes, 2No	Viet, Thai
1	12	1M	Pupil	1 No	Japanese
Total: 30					

database.

3.3 Materials and Method

3.3.1 Participants

The database contains 30 subjects from 11 year-old to 32 year-old as depicted in table 3.4. To ensure accuracy in results of the experiments, all of the participants were asked to take rest, maintain in good mood for 2 hours prior to the measurements and to avoid the presence of cosmetic substances on their face at the time of experiment. Participants have not had cardiovascular or respiratory conditions and are not taking medications at the time of the experiments. They allowed us of their visible-spectrum and infrared images in scholarly publications. Before taking the experiment, each participant consented to join the test and also signed the test agreement. At the beginning of each data acquisition, they are explained about the object of the experiments, methods, procedures, and potential benefits of the experiments. Participant are informed about ethical experiment design practices and protocols. During each experiment, the participants could quit any stages if they do not want to join it. At the end of each stage of data acquisition, participants must give self-reports designed by us.

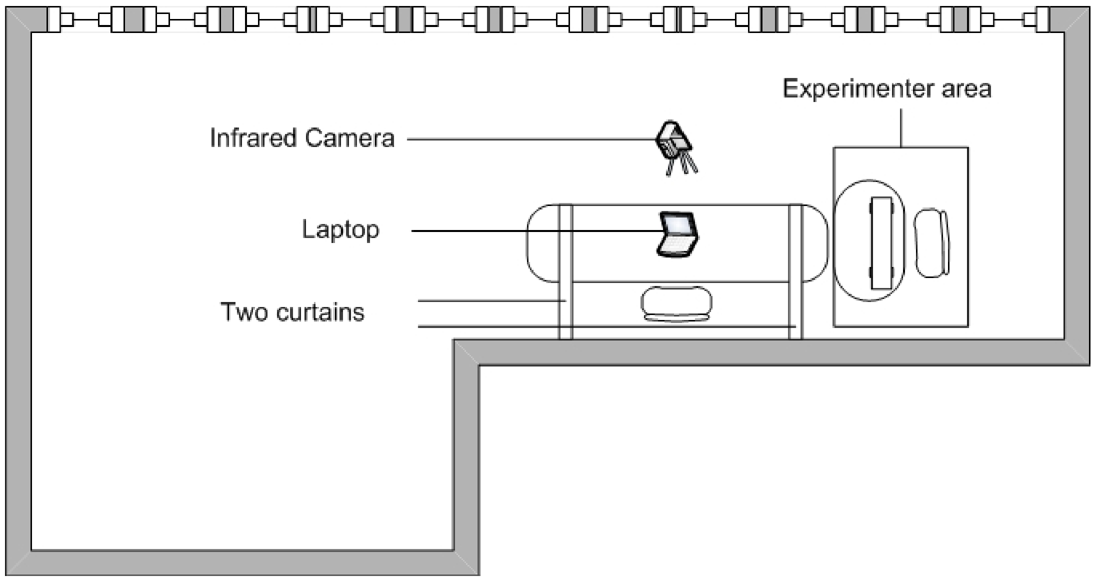


Figure 3.2: Overview of experiment room.

3.3.2 Measurement Devices and Environment

Room Setup

The room for conducting the experiment is L shaped with 8m∗12m ∗3.5m and the omitted area is about 6m^2. The experiment room was always kept quiet to ensure no effect to induce the participants emotions. During the data acquisition, the internal temperature of the room that is used for conducting experiments is maintained between 24°C and 26°C because of the sensitivity of the facial surface to the environmental temperature. To control the humidity and temperature of the room, we used the building air conditioning system, and the flow of air condition was not directed to the testing area. To keep the constant illumination between day and night, kept both the door and the curtains closed during the experiment. The experiment area had infrared camera, equipped with laptop, desk, chair, LCD screen, mass storage disk, head phone, and two special curtains. Two curtains separated the participant from the experimenter as a result of which the participants felt more comfortable and not shy, making it easier to induce their emotions. The view of the room and experiment area is depicted in Fig.3.2.

Camera Setup

We used an Infrared Camera NEC R300 to obtain the visible and thermal IR videos. The general construction and introduction of user guide for R300 are introduced in Figure 3.4, 3.5, 3.6,3.7, 3.8, Table 3.5, 3.6. The infrared camera has 3.1mega pixels visible camera capturing 5 frames per second and a long wavelength infrared (LWIR) camera opening from 8μm to 14μm. The thermal sensitivity is 0.03°C at 30°C. Thermal infrared imaging data were captured at 5ft/s. The camera was placed at a height of 1.5m above the floor and 0.85m in front of the participants. To obtain the correct temperature of

No.	Items	Description
1	Status icon	State of camera is displayed by the following icons. **<Display of status icons>** Battery life : Indicates remaining battery in four stages →P.21 SD Card : Appears when SD card is inserted →P.1-3 Alarm execution : Appears when alarm is raised →P.3-10、P.4-41 Averaging execution : Appears while averaging is executed →P.4-39 Ambient compensation : Appears while distance correction is executed →P.4-32 Temperature value setting: Appears when temperature scale is set to Serial Auto →P.2-2 Freeze : Indicates freeze status →P.2-1 REC :Appears while saving data into SD memory card →P.4-3, P.4-4, P.4-6 Lens assembly : Appears when lens correction is executed →P.4-34 Save mode : Indicates current save mode →P.4-2, P.4-4, P.4-6 Trigger save mode : Appears when the external event trigger save setting is enabled → P.4-45 USB image transfer mode: Appears when USB image transfer mode is set → P.4-62

Table 3.5: Overall user guide of R300.

No.	Items	Description
2	Alarm message & error message	Below messages are displayed when alarm condition is satisfied (Background: yellow, Text: red) (1) "ALARM" :Displayed when the alarm generation conditions in the alarm settings are met. (2) "BATTERY" :Displayed when the battery is low. Error message Displayed when the error conditions are met. (Background: red, Text: black) (1) " MEMORY" :Displayed when Internal data is not normal. (Data of the unit may not be accurate.) →Page 5-2 (2)STABILIZE :Displayed when the temperature sensor is not stabilized. (3)FILTER :Displayed when the filter motor operation is not normal. (4)FOCUS :Displayed when the focus motor operation is not normal. (5)TEMP S/TEMP L/TEMP R: Displayed when either of the temperature sensors (shutter/lens/Rad) encounters a read error. (6)VIS COM :Displayed when a visible camera communication error (I2C) is encountered. (7)LENS COM :Displayed when a communication error with the lens unit (UART) is encountered. (8)SD ACCESS :Displayed when an SD card access error is encountered. (9)BAK MEM :Displayed when a camera setting value (backup value data) is abnormal.
3	Date and time	It displays date and time. * Displayed clock type is changeable. →Page 1-4
4	Range	(1)60℃ :Range1 (-20～60℃) (2)120℃ :Range2 (-40～120℃) (3)500℃ :Range3 (0～500℃) (4)2000℃:Range4 (200～2000℃) *Range4 is option
5	Upper temperature limit	It displays the upper temperature limit of color bar. * Upper temperature limit can be changed. →P.2-1, P.3-7, P.4-13
6	Lower temperature limit	It displays the lower temperature limit of color bar. * Lower temperature limit can be changed. →P.2-1, P.3-7, P.4-13
7	Color alarm upper temperature limit	It displays the upper temperature limit of color alarm.
8	Color alarm lower temperature limit	It displays the lower temperature limit of color alarm.
9	Color bar	It displays color bar.
10	Unit of temperature value	It displays temperature unit.
11	Temperature value at BOX	Indicates the maximum, minimum and average temperatures in specified box area (up to five areas).
12	Temperature value at point cursor	Indicates a temperature of specified point (up to 10 points).
13	Max/Min temperature value	Indicates a temperature of trace cursor. Max: Maximum temperature within a specified area min: Minimum temperature within a specified area
14	ΔT temperature value	Indicates temperature difference between two specified points. [Display example] M-a: Maximum temperature(Max) – point a a-m: Point a – minimum temperature (min)
15	Emissivity value	Indicates a preset emissivity value. * This can be changed. →P.3-11, P.4-30
16	Ambient temperature value	Indicates a temperature around the camera.

Table 3.6: Overall user guide of R300.

Figure 3.3: InfReC thermography camera NEC R300.

participants, the calibration was set up before each experiment and updated automatically per minute. We used NS9500 PRO, supporting real-time monitor, to capture both visible and thermal IR data and NS9500 STD to view, enhance, analyze, and extract the thermal IR data.

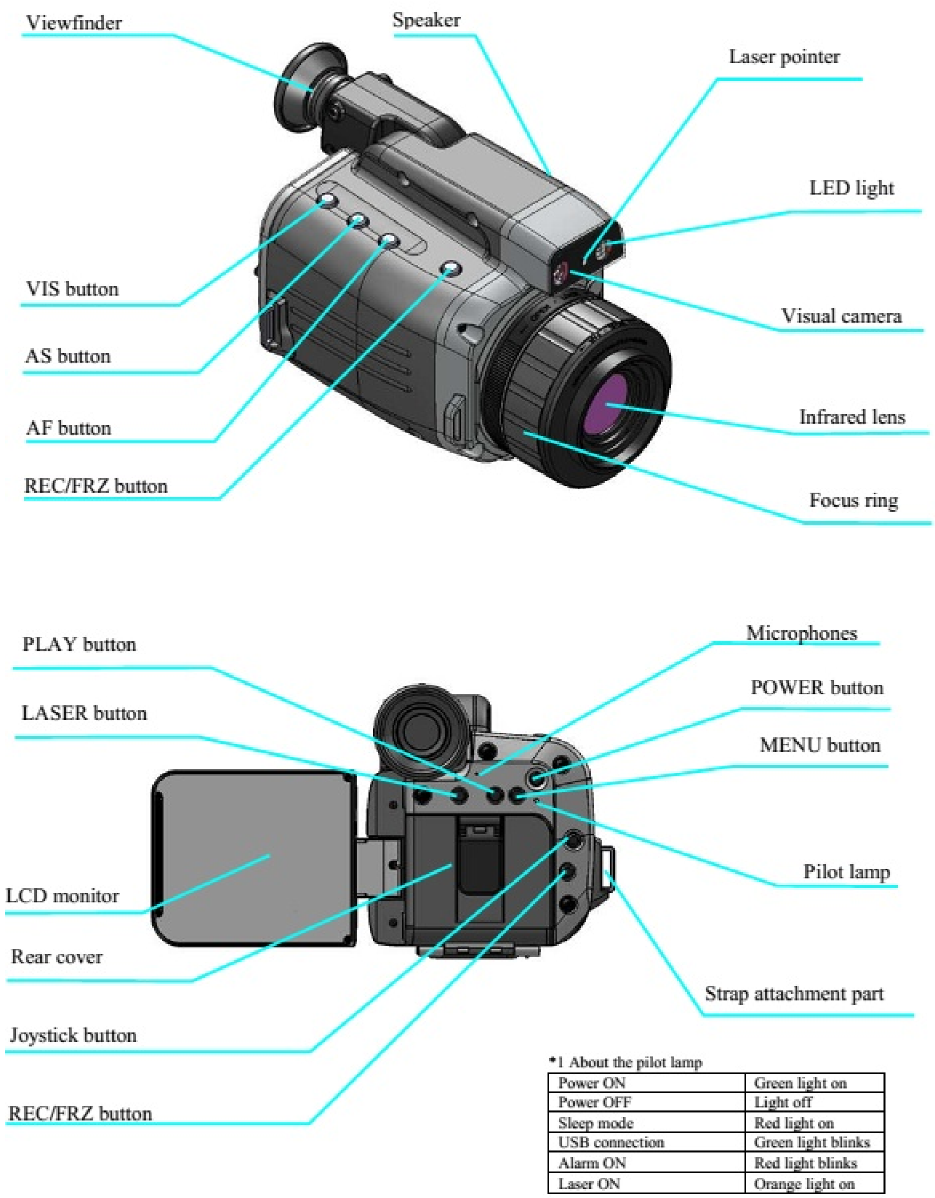
3.3.3 Procedures

Stimuli

In this experiment, we use selected emotional video clips to evoke the emotion of the participant. The video clips were gained by four persons from the internet and judged by the authors. There are four angry clips, four disgust clips, four fearful clips and one fearful game, six happy clips, seven sad clips, three surprised clips and two neutral clips. We further classified each emotion class into four sub-classes according to their intensive levels.

Data Acquisition

The experiment room had only one participant and experimenter during data collection. The participant was seated comfortably in an armchair in front of a laptop screen as shown in Fig.3.2. Fig.3.9 shows the data acquisition procedure, which fixes the former database mistake. Depending on the participants, we did not ask them to not wear on or take off their glasses. We also did not require them to keep their head fixed in one position because we wanted to obtain the spontaneous emotions. The participants were given an



Power ON	Green light on
Power OFF	Light off
Sleep mode	Red light on
USB connection	Green light blinks
Alarm ON	Red light blinks
Laser ON	Orange light on

Figure 3.4: Overall construction of R300.

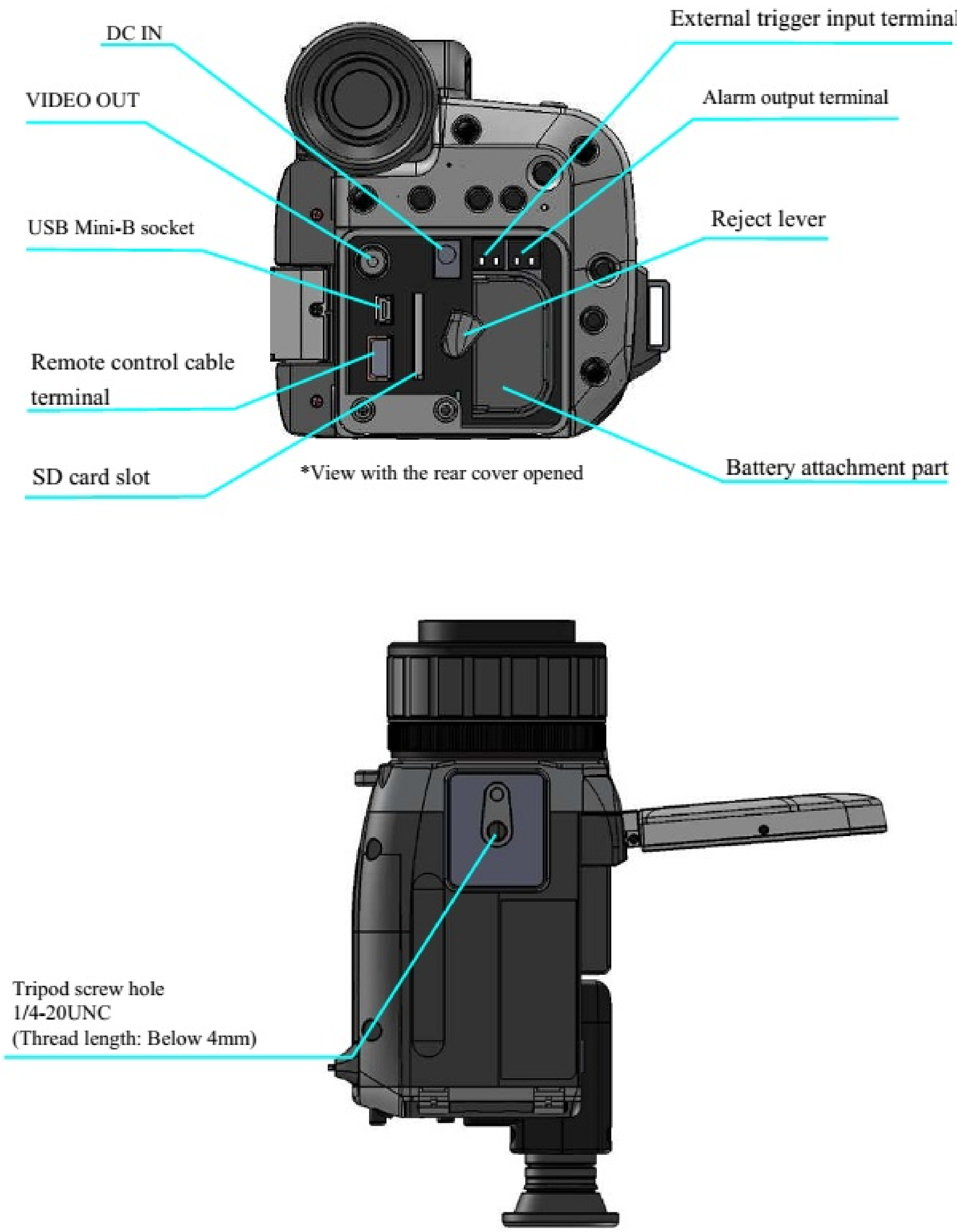
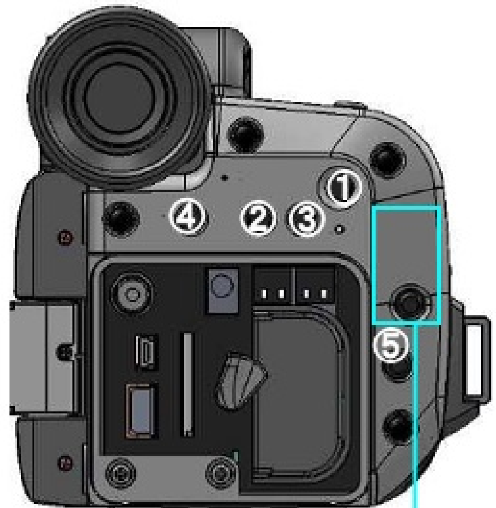


Figure 3.5: Overall construction of R300.

<Button>



(1) POWER button
It turns ON or OFF the power.

(2) PLAY button
It makes the unit in image replay mode.

(3) MENU/CANCEL button
For displaying menu.

(4) LASER button
Laser pointer is projected while holding down this button.

(5) REC/FRZ button
Press briefly: Switches between loading of thermal images (run) and stop (freeze).
Press and hold: Saves thermal image.

(6) Joystick button

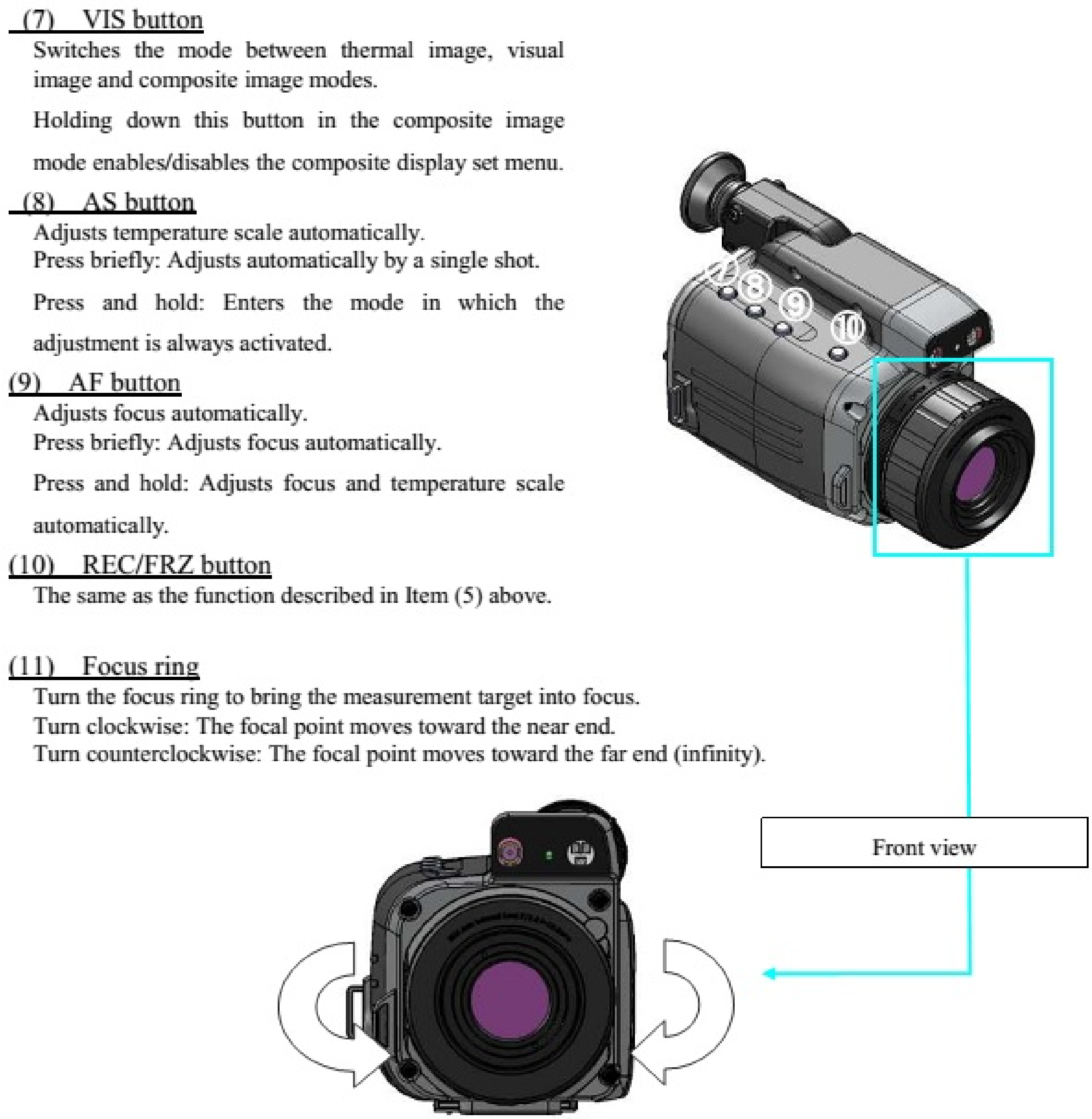
H Temp (UP):	For setting the maximum value of the temperature scale. It also works as a direction key (UP).
L Temp(DOWN):	For setting the minimum value of the temperature scale. It also works as a direction key (DOWN).
NEAR(LEFT):	The focus point moves toward the near end. It also works as a direction key (LEFT).
FAR(RIGHT):	The focus point moves toward the far end (∞). It also works as a direction key (RIGHT).
ENTER:	Press briefly: Enters the level/span setting mode (*1). Press and hold: Displays the short cut menu.

*1 Level/span setting mode:
Level: Sets a displayed center temperature (level) so that a measured temperature is near the center temperature when displaying a thermal signal obtained from the sensing section.
Span: Sets a measurement sensitivity (span) when displaying a thermal signal obtained from the sensing section.

The button functions in each of the setting modes are as shown below.

	Level/span setting mode	Maximum/minimum value of the temperature scale setting mode
H Temp (UP)	Increases the level.	Increases the selected value.
L Temp (DOWN)	Decreases the level.	Decreases the selected level.
FAR (RIGHT)	Increases the span.	Moves the digit to the right.
NEAR (LEFT)	Increases the span.	Moves the digit to the left.
ENTER	Confirms the setting value to reflect it.	Confirms the setting value to reflect it.

Figure 3.6: Overall user guide of R300.



(7) VIS button

Switches the mode between thermal image, visual image and composite image modes.

Holding down this button in the composite image mode enables/disables the composite display set menu.

(8) AS button

Adjusts temperature scale automatically.
Press briefly: Adjusts automatically by a single shot.

Press and hold: Enters the mode in which the adjustment is always activated.

(9) AF button

Adjusts focus automatically.
Press briefly: Adjusts focus automatically.

Press and hold: Adjusts focus and temperature scale automatically.

(10) REC/FRZ button

The same as the function described in Item (5) above.

(11) Focus ring

Turn the focus ring to bring the measurement target into focus.
Turn clockwise: The focal point moves toward the near end.
Turn counterclockwise: The focal point moves toward the far end (infinity).

Figure 3.7: Overall user guide of R300.

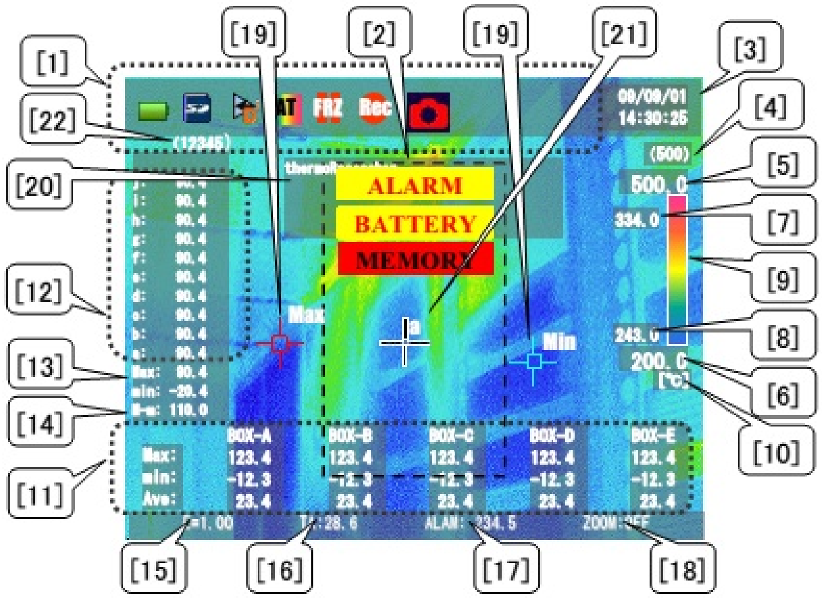


Figure 3.8: Overall user guide of R300.

introduction of the purpose and procedure of the experiment prior to taking data and then they were asked to wear headphone. Before and after each session, the instrumental music turned on to help the participant return to the neutral feeling. In each session, we tested only one emotion and specially paid attention to the time lag phenomenon. When participants did not to want to watch fearful or angry clips, we stopped the session to respect their right. After experiment of each person, we asked them to give the report contributed their feeling to each emotion video clips. These self-reported data were helped us to label the recorded videos.

3.3.4 KTFE Database Design

The first version of KTFE database includes 186 gigabyte of visible and thermal IR facial emotion videos, visible facial expression image database and thermal facial expression image database. To obtain the thermal expression image database, we manually choose the expressions using NS9500STD software. A team of three people selects manually the suitable frames for every emotion of each person and extract into thermal images. The visible image database is also manually extracted and chosen by two people.

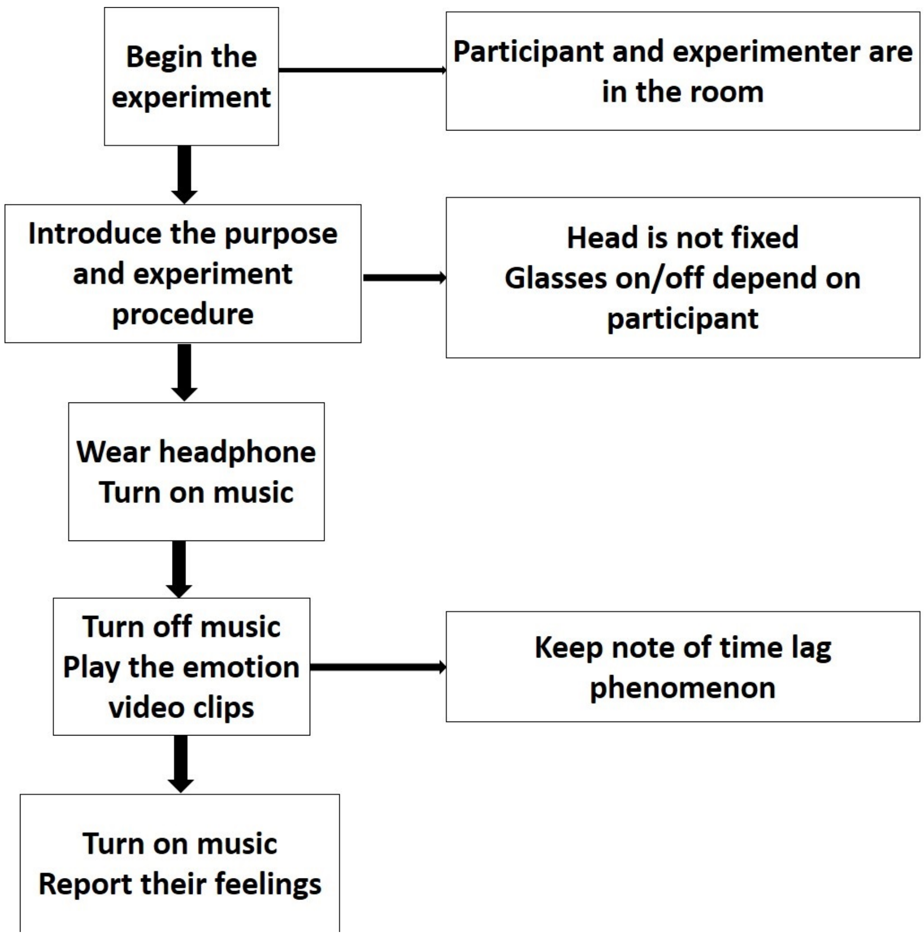


Figure 3.9: Data acquisition procedure.

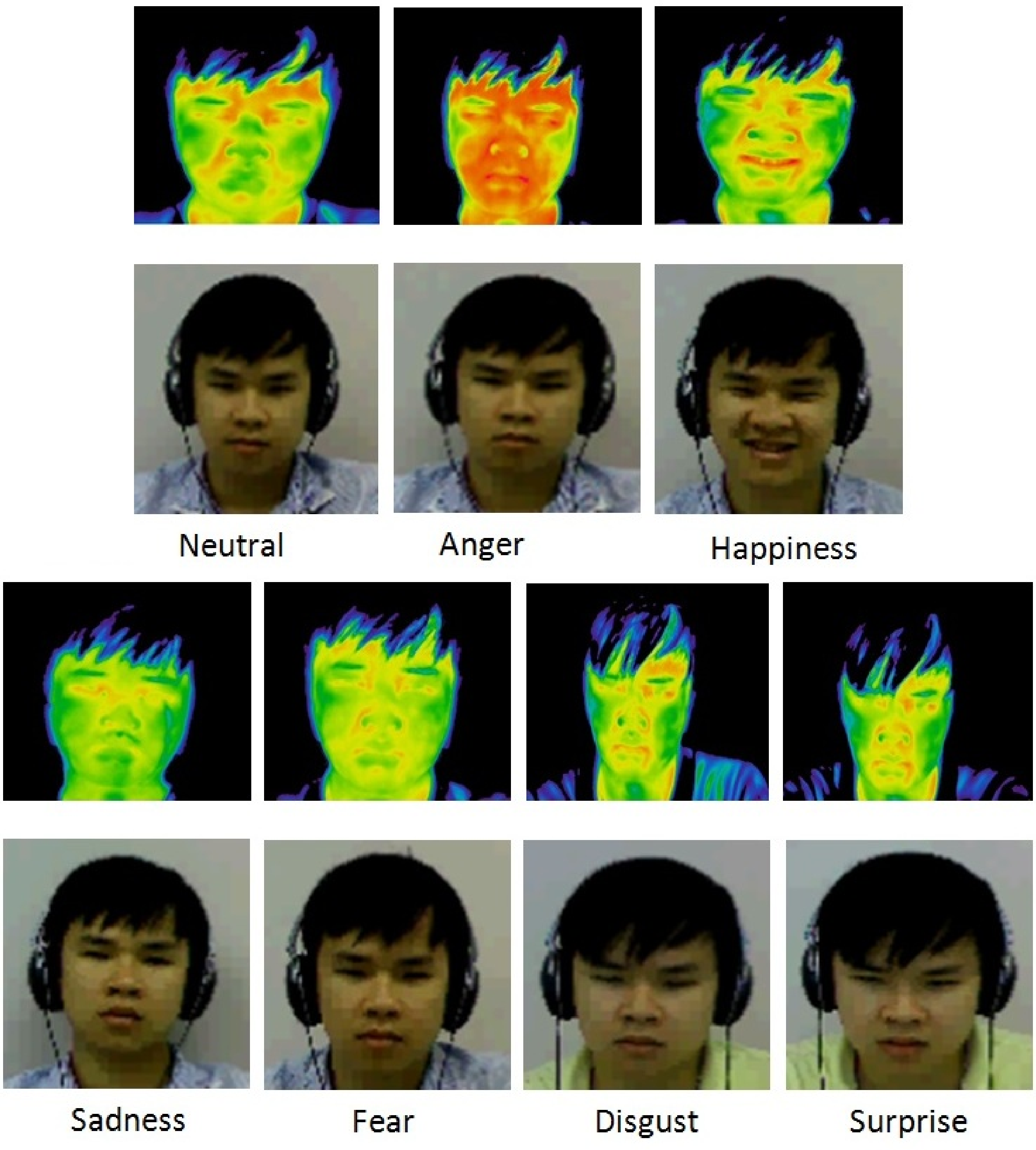


Figure 3.10: Sample thermal IR and visible images of seven expressions.

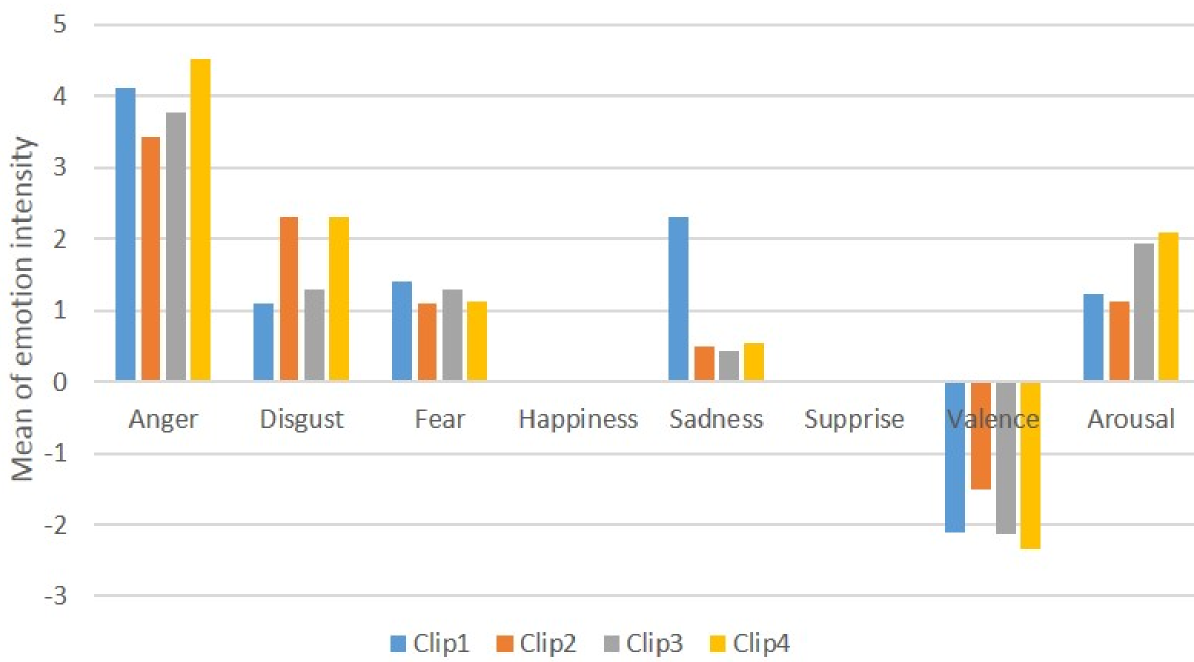


Figure 3.11: Video used to induce anger.

3.4 Data Analysis

3.5 Analysis of the Effectiveness of Eliciting Video

In this section, we evaluate a method of emotions elicitation. Using self-report of participants about their emotions, we estimate the effectiveness of emotion-elicitation video clips.

3.5.1 Methodology

In the database acquisition process, after each process, participants will make a self-report about their feeling. In each self-report, they will answer some questions which are designed by us. For example, "Have you ever watched this clips?", "What are your feelings (picking up from 7 emotions)?", " In 5-levels of feeling, which scale are you feeling? and so on. From the most view clips, we choose four clips for sadness, four clips for fear, fours clips for disgust, four clips for happiness, four clips for sadness and three clips for surprise. We calculate the mean evaluation values of each emotions as well as the valence and arousal, which reflect the overall evaluation results for each emotion-eliciting video clips [31].

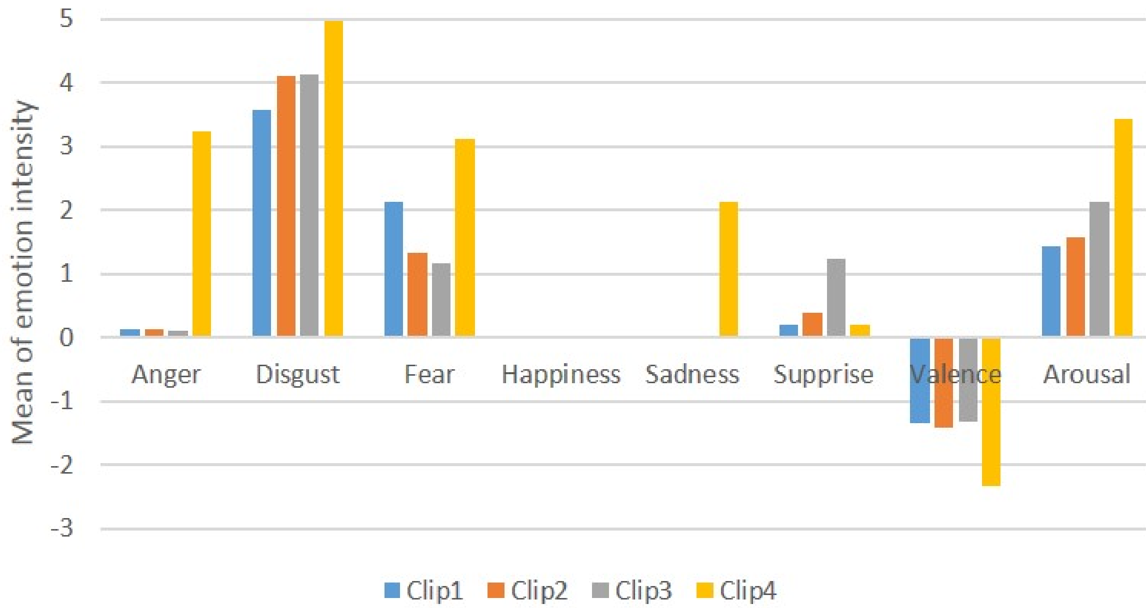


Figure 3.12: Video used to induce disgust.

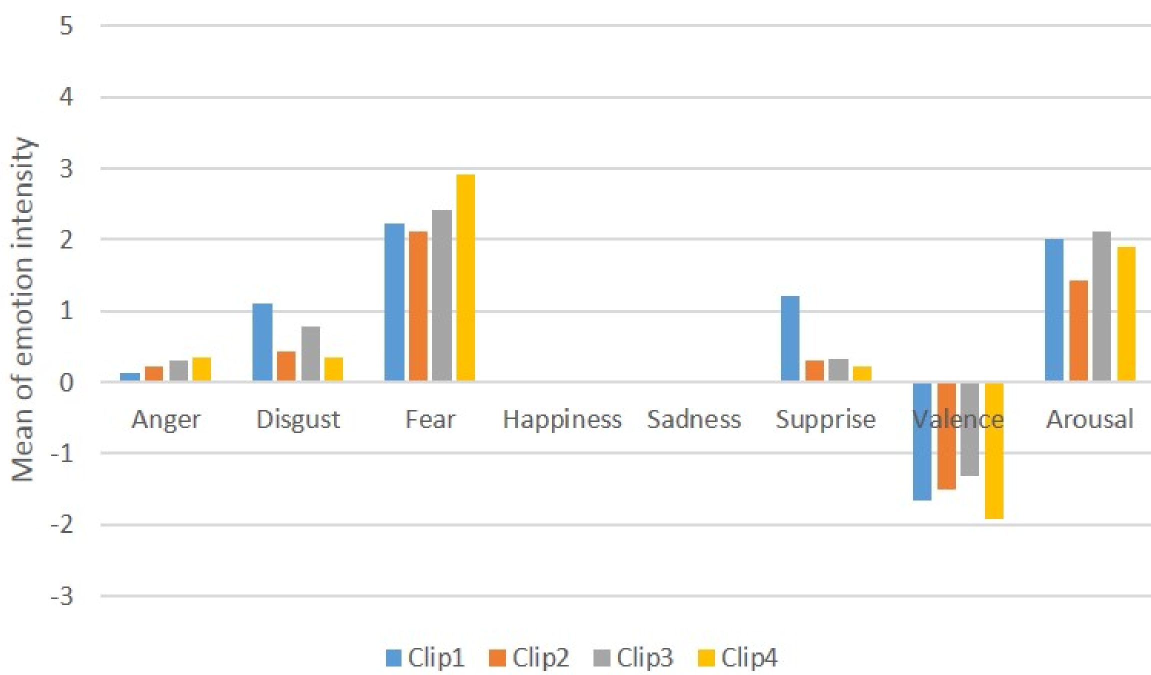


Figure 3.13: Video used to induce fear.

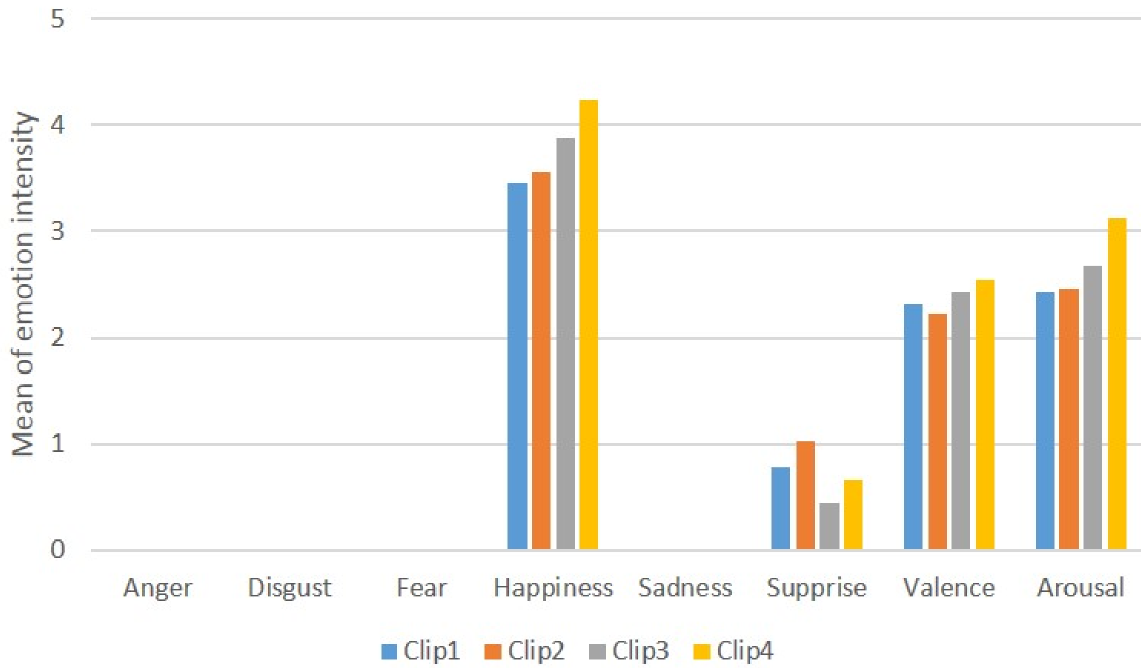


Figure 3.14: Video used to induce happiness.

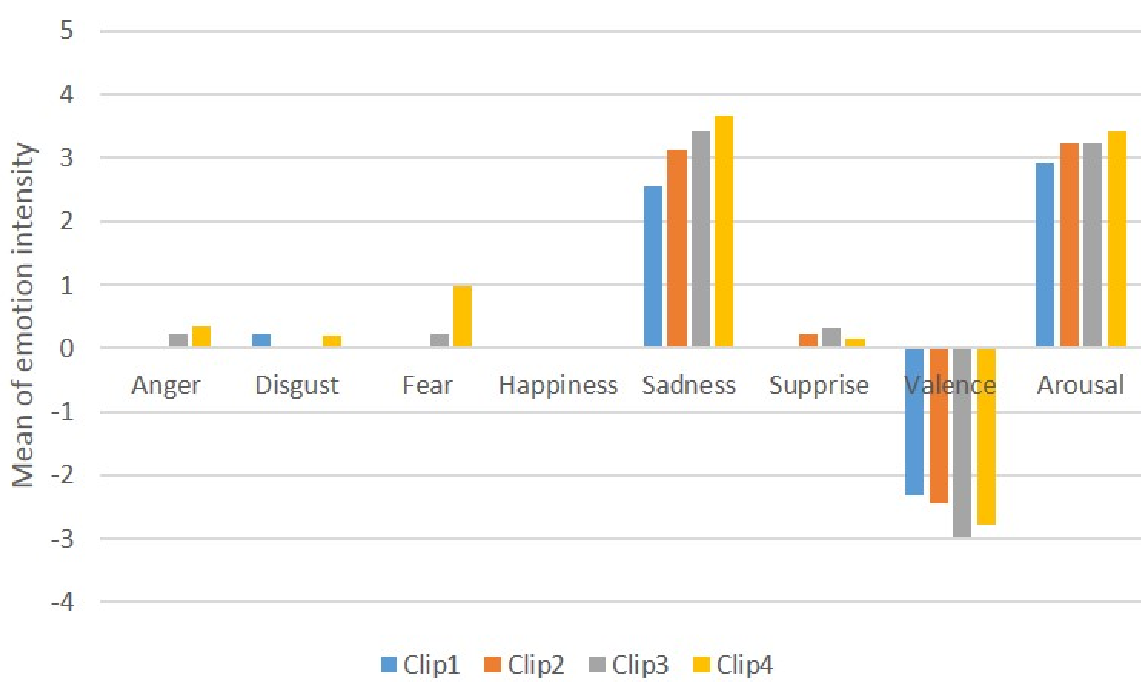


Figure 3.15: Video used to induce sadness.

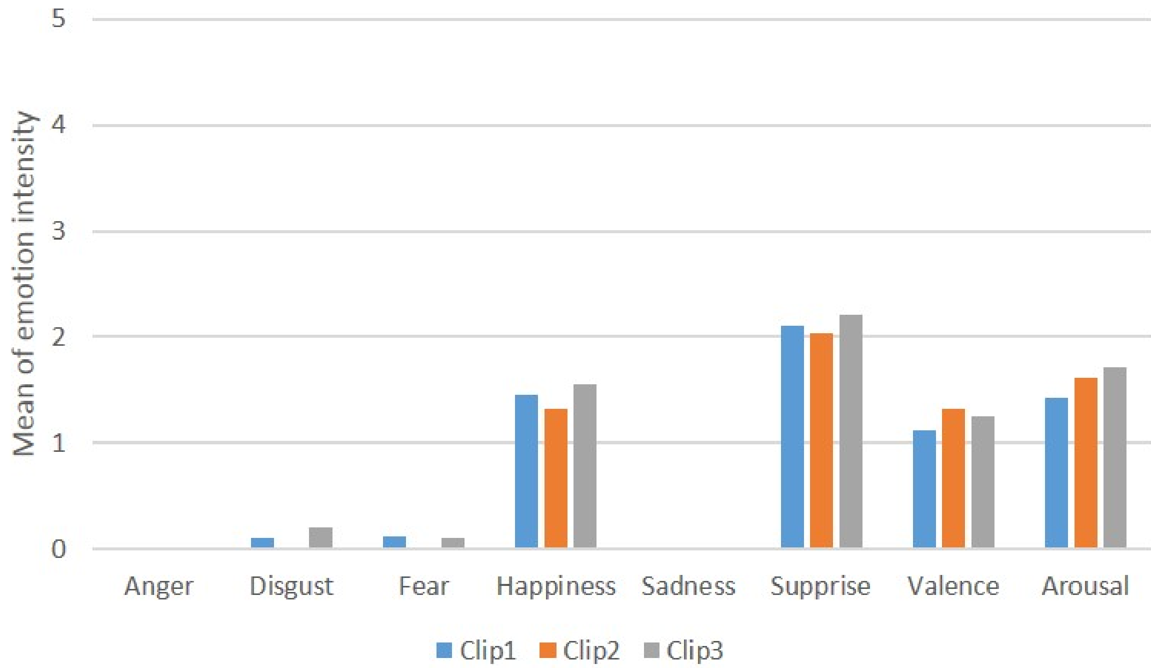


Figure 3.16: Video used to induce surprise.

3.5.2 Results and Analysis

From figure 3.11 to figure 3.16, the mean of participant self-report data for each emotion are shown. We have some conclusion from those results:

Firstly, according to those pictures, most of the video clips are induced the desire emotions. Therefore, our clips almost work well and effectively. Another cue to support the effectiveness of our clips is the means of the valences. With happiness, surprise, the positive emotions have positive means of the valences, and the negative emotions, sadness, anger, disgust, fear have negative means of the valences.

Secondly, the positive means of arousal prove that all video clips are induced emotion almost successful.

From figure 3.11 to figure 3.16, we can infer that the video clips used to evoke a emotion could induce multiple emotions. For example, the clips for anger can evoke some degree of the disgust, fear and sad emotions. The clips for disgust can indicate some degree of the anger, fear and surprise. The clips for happiness can evoke some degree of surprise. This is consistent with the previous study results described in [31], [120]. From those phenomenon, we can develop to new complex emotions which are not limited to six basic categories. They also give some prior information to support for estimate human emotion.

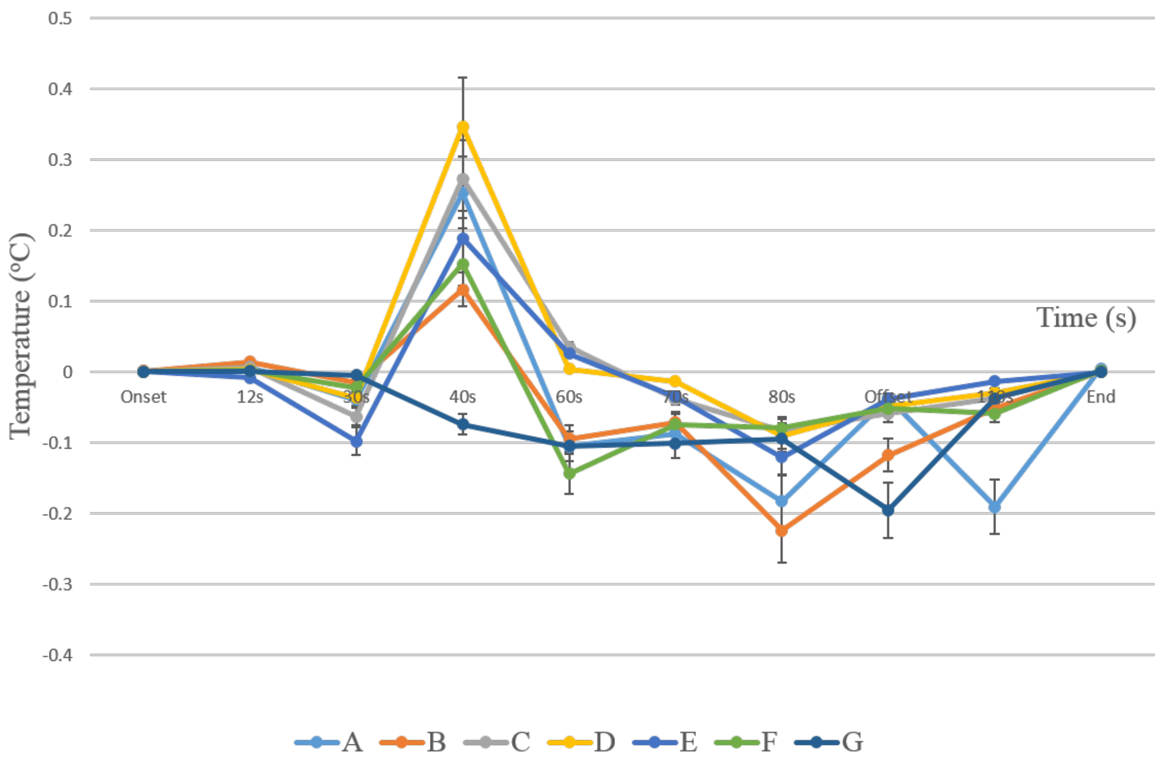


Figure 3.17: Temperature change of happiness.

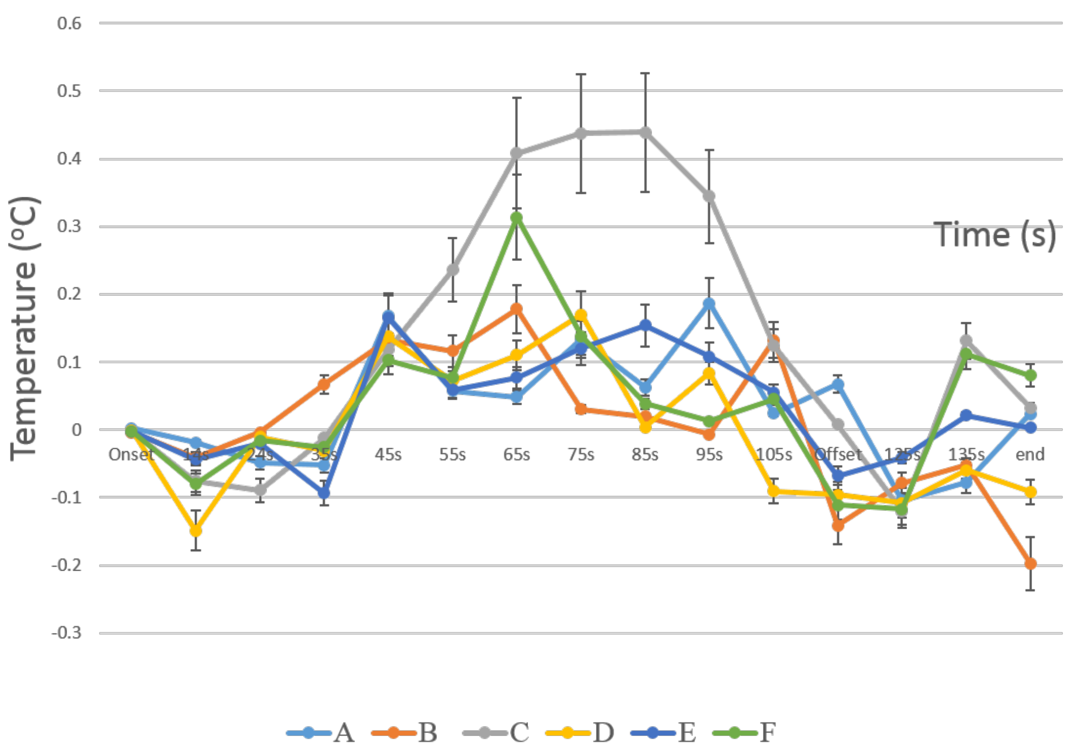


Figure 3.18: Temperature change of sadness(crying).

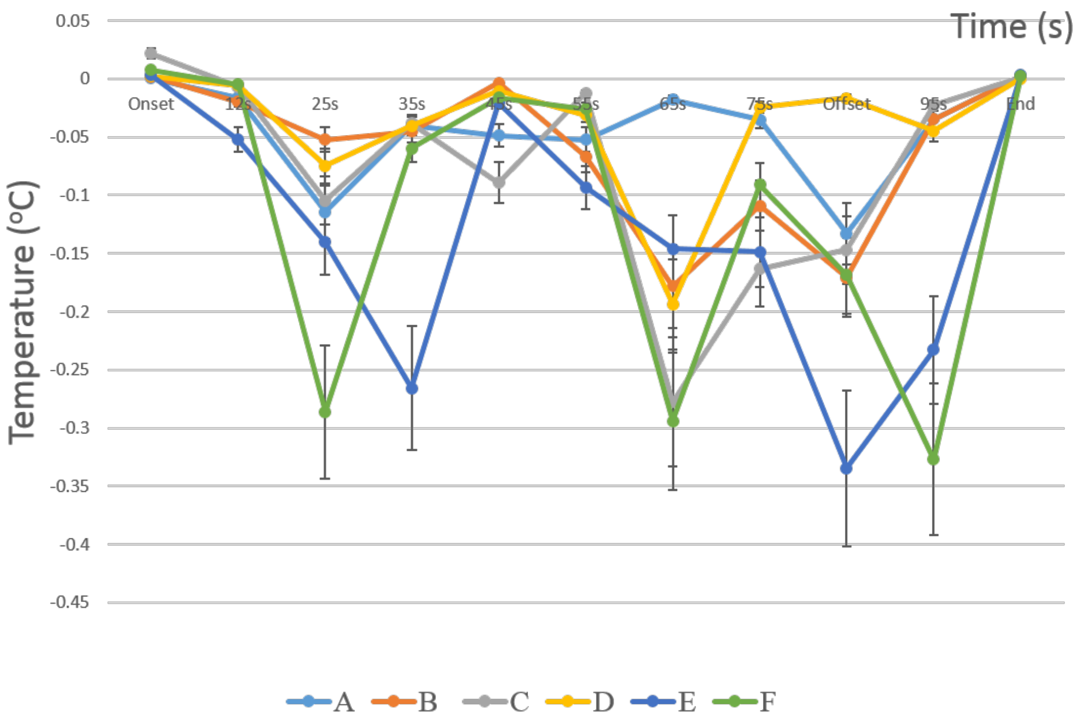


Figure 3.19: Temperature change of sadness.

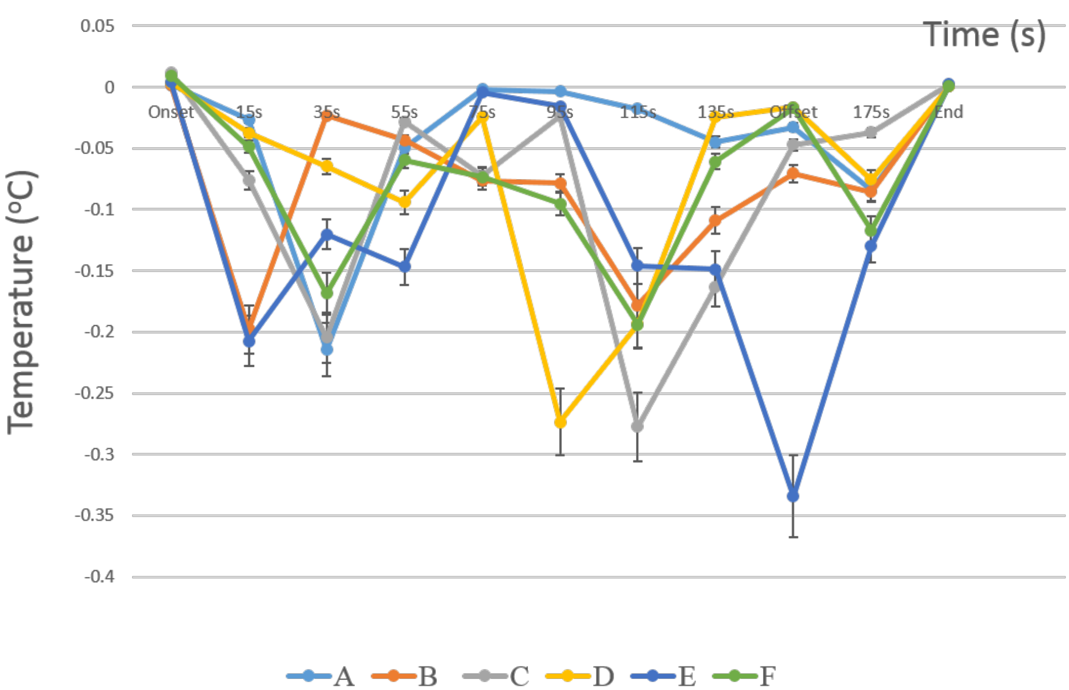


Figure 3.20: Temperature change of disgust.

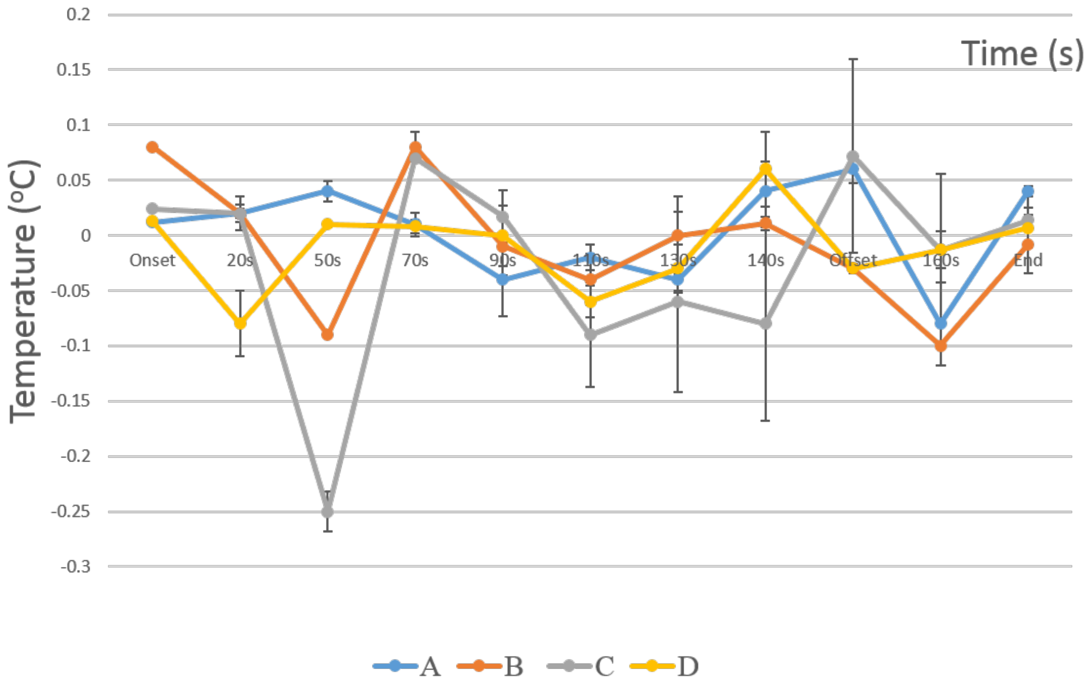


Figure 3.21: Temperature change of surprise.

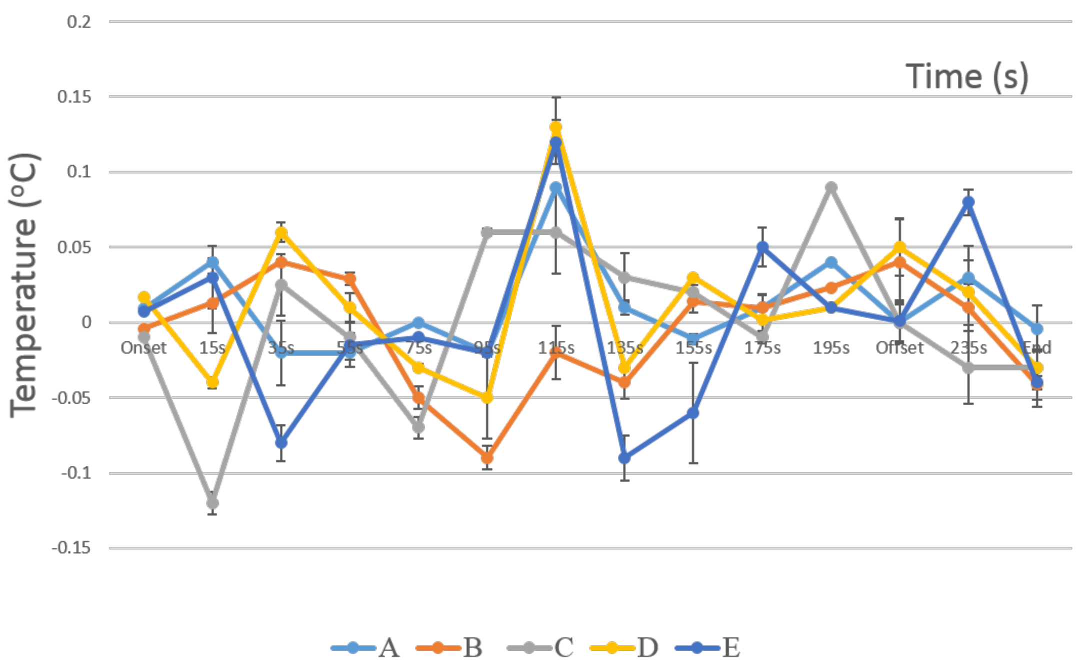


Figure 3.22: Temperature change of fear.

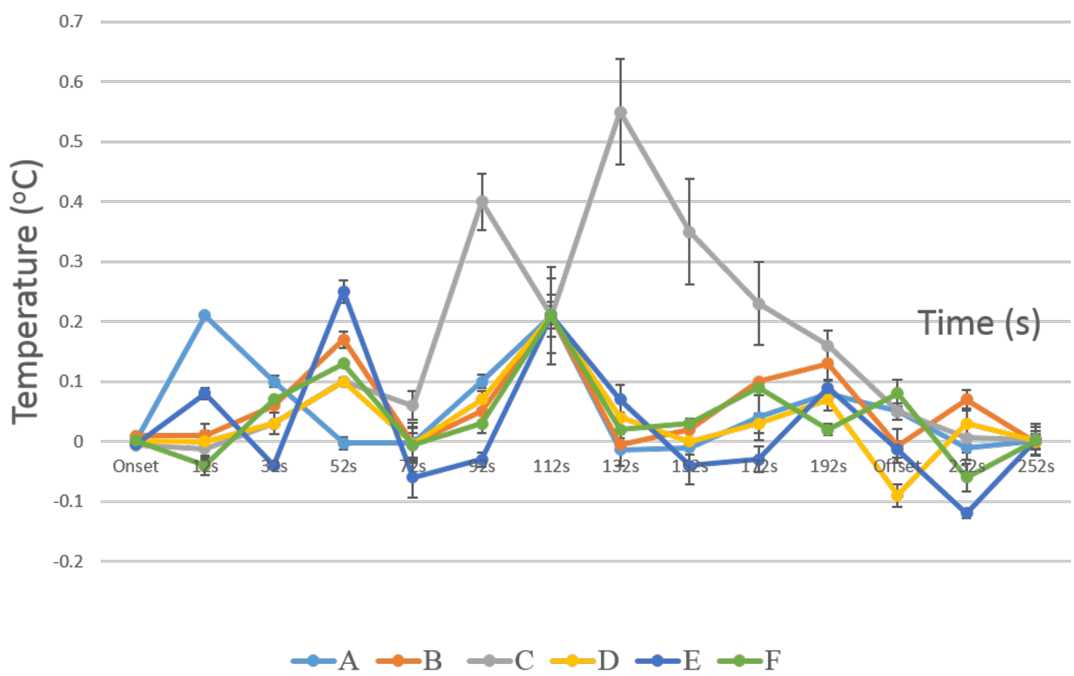


Figure 3.23: Temperature change of anger.

3.6 Analysis of Temperature Change

We calculate the mean of all participant's facial area temperature of each emotion from onset to offset of emotion. The facial area are interest regions of face where temperature increases and decreases significantly when emotion changes. Here, we use one to three areas for forehead, two areas for cheek-bone, one to two for noses.

Based on figure 3.17, with A,B and C part are in forehead, D and E part are in eye-holes, F part is in nose, G part is in cheek-bone, we can extract some interesting information about happiness. From onset to offset, it takes approximately 10s to change temperature. Most of the areas will increase the temperature of the first phrase, and decrease for others remaining phrases. Only the cheek-bone left temperature is always decrease.

Using figure 3.18 and figure 3.19, with A part is in forehead, B part is in eye-holes, C part is in nose, D and E part are in cheek-bone, F part is around nose, the temperature of sadness separates two kind of changing, firstly for crying, after time-lag approximately 10s, at the first phrase, temperature decreases, and then others remaining phrases, temperature of all chose facial parts decrease, secondly for normal sadness, after time-lag approximately 10s, temperature decreases for all phrases.

From figure 3.20, with A and B part are in forehead, C part is in eye-holes, F part is in nose, D and E part are in cheek-bone,temperature of disgust in all areas for all phrases decreases after time-lag approximately 10s.

Following figure 3.21, with A part is in forehead, B and C part are in cheek-bone, D part is in nose, temperature of all phrase has no changing rule. However, based on total of temperature means, the trend of temperature is down.

Following figure 3.22, with A part is in forehead, F part is in cheek-bone, D part is in nose, E part is around mouth, B and C part are in eye-holes temperature of all phrase has no changing rule. However, based on total of temperature means, the trend of temperature is down.

Following figure 3.23, with A and B are in forehead, C part is in eye-holes, F part is in nose, D and E part are in cheek-bone, for anger emotion, after approximately 10s from the onset, temperature of two areas of forehead, cheek-bone, nose increases. This information prove that with anger emotion, the blood pressure increases than temperature face increases.

In conclusion, for those information, the temperature changes of happy, sad, anger, disgust emotion are following some rules, however temperature changes of fear and surprise from our database are not following any rules. Using the obtained rules, we can use to support for estimating the human emotions. Those are the very important prior information to support for estimating human emotions.

3.6.1 Estimation of Human Emotion

In this section, we bring a fundamental evaluation of usability of the visible facial database, thermal facial database and conduct the thermal IR data to analyze emotions. Before evaluating of these databases, to avoid any undesired noise in the thermal IR images, visible images, we use median smoothing filter to reduce noises and blurring. The preprocessing for visible image is to normalize image to facial space containing only face. To analyze the facial expression using thermal IR data of expressions, after reducing the effect of noise, we use three convention methods PCA, EMC, PCA-EMC. With PCA, the aim is to build a face space, including the basis vectors called principal components, which better describe the face images [113]. To estimate emotions using PCA, we divide the training set into five classes and compute the eigen-space as following:
Step 1: Concatenating each row of a thermal IR data of each frame by row, the thermal IR data can be transformed to a column vector. Given M frames of thermal IR data as training data, we convert these datum to corresponding column vectors
Step 2: The mean of training data has to be calculated and then subtracted from each original thermal IR data in the datum.
Step 3: Calculate the eigenvectors and eigen-values of the covariant matrix.
For each test thermal IR data, we project it to the eigen-space of each class and derive the reconstruction thermal IR data from each eigen-space. Using mean square error, measuring the similarity, between input thermal IR data and reconstruction thermal IR data, we can choose a suitable class for input thermal IR data which is a minimum of mean square errors [113].

In Reference [114], the authors proposed eigen-space method based on class-features (EMC) to analyze the facial expressions. The difference between PCA and EMC is that PCA finds the eigenvector to maximize the total variance of the projection to line, while EMC is obtained eigenvector to maximize the difference between the within-class and between-class variance. The Fig.3.24 shows the advantage of EMC over PCA. The difference between the within-class and between-class covariance is calculated as following:

$$S = S_B - S_W. \tag{3.1}$$

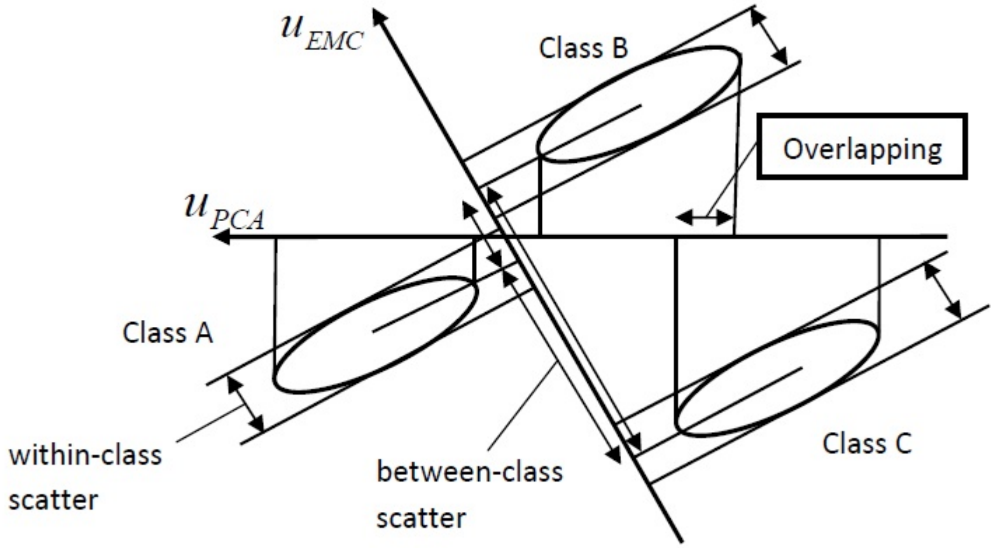


Figure 3.24: Examples of a eigenvector of PCA and EMC [115].

$$S_B = \frac{1}{M}\sum_{f\in F} M_f(\overline{x}_f - \overline{x})(\overline{x}_f - \overline{x})^\tau. \tag{3.2}$$

$$S_W = \frac{1}{M}\sum_{f\in F}\sum_{f\in F}^{M_f} M_f(\overline{x}_{fm} - \overline{x}_f)(\overline{x}_{fm} - \overline{x}_f)^\tau. \tag{3.3}$$

$$\overline{x}_f = \frac{1}{M}\sum_{m=1}^{M_f} x_{fm}; \overline{x} = \frac{1}{M}\sum_{f\in F}\sum_{m=1}^{M_f} x_{fm}. \tag{3.4}$$

where F is a set of expression classes, M_f facial-patterns are given for each class $f \in F$ and x_{fm} is an N-dimension vector of the $m - th$ facial patterns,$m = \overline{1, M_f}$

To estimate emotion using EMC, we divide the training set into five classes and compute the eigen-space. For each test thermal IR data, we project it into the eigen-space of each class and an emotion is chosen if it gets maximum of cosine of angle between obtained vector after projection and eigenvector of each class. With PCA-EMC, we use PCA to reduce the dimension and then apply EMC to the obtain data.

3.6.2 Evaluation of the Visible Image Database

The preprocessing for visible image is to normalize image to facial space containing only face. The three well-know algorithms are used to classify images to emotions. There are PCA [113], EMC[114] and PCA-EMC methods. We extracted the visible image database from KTFE database. It includes 330 images of 22 subjects for 5 expressions. The rate of testing and training is 40 percent and 60 percent of total images, respectively. Table 3.7, Table 3.8, Table 3.9, including confusion matrices and average recognitions, shows the

Table 3.7: Confusion matrix of expression analysis of visible images with PCA.

	Ag	Ha	Fe	Ne	Sa
Ag	78.23	6.06	7.51	8.20	-
Ha	2.30	46.62	8.35	40.08	2.65
Fe	5.46	5.39	82.80	6.35	-
Ne	5.96	34.79	5.92	36.31	17.02
Sa	-	12.68	3.65	14.09	69.58
Avg	62.71				

Table 3.8: Confusion matrix of expression analysis of visible images with PCA-EMC.

	Ag	Ha	Fe	Ne	Sa
Ag	84.42	5.00	9.33	1.25	-
Ha	6.73	51.80	2.67	36.26	2.54
Fe	3.30	2.69	86.23	5.83	1.95
Ne	6.78	32.73	7.05	42.62	10.82
Sa	-	3.83	0.67	5.71	89.79
Avg	70.97				

Table 3.9: Confusion matrices of expression analysis of visible images with EMC.

	Ag	Ha	Fe	Ne	Sa
Ag	71.88	6.05	8.84	13.23	-
Ha	5.04	42.36	7.76	41.03	3.81
Fe	0.98	9.22	74.51	12.35	2.94
Ne	9.08	39.44	8.58	28.13	14.77
Sa	-	7.08	1.76	13.7	77.46
Avg	58.87				

performance of these algorithms with respect to our visible image database. According to confusion matrices, anger expression do not recognize by sadness and vice versa. There are some confusion between happiness and neutral because some persons do not reveal their smiling.

3.6.3 Evaluation of the Thermal IR Data

To classify emissions using thermal data, we used PCA, PCA-EMC. The thermal data was extracted from KTFE. It includes 3.5GB thermal data for five emotions. The training and testing data are 60 percent and 40 percent of the total thermal data. From table 3.10, 3.11, 3.12, no instance of anger is incorrectly recognized as happiness. Neutral emotion and anger emotion are the most recognized by PCA, and PCA-EMC and EMC, respectively. Based on the results, we confirm that thermal data are important addition information to support for expressions and emotions analysis.

Table 3.10: Confusion matrix of emotion classification of thermal IR data with PCA.

	Ag	Ha	Fe	Ne	Sa
Ag	59.17	0.00	22.50	0.00	18.33
Ha	0.00	61.86	17.61	6.80	13.73
Fe	9.92	13.58	57.34	1.54	17.62
Ne	10.19	9.51	3.08	69.82	7.40
Sa	5.96	22.21	2.35	4.35	65.12
Avg	62.66				

Table 3.11: Confusion matrix of emotion classification of thermal IR data with PCA-EMC.

	Ag	Ha	Fe	Ne	Sa
Ag	81.33	0.00	18.67	0.00	0.00
Ha	0.00	65.98	10.49	10.76	12.77
Fe	9.60	11.46	54.35	7.83	16.76
Ne	17.01	16.05	5.17	56.95	4.82
Sa	8.31	14.07	2.23	7.79	67.60
Avg	65.24				

Table 3.12: Confusion matrix of emotion classification of thermal IR data with EMC.

	Ag	Ha	Fe	Ne	Sa
Ag	81.33	0.00	14.67	4.00	0.00
Ha	2.71	64.64	12.03	4.12	16.50
Fe	5.88	12.28	55.60	1.18	25.06
Ne	5.17	10.67	9.97	58.29	15.90
Sa	1.36	11.26	12.83	10.26	64.29
Avg	64.83				

Table 3.13: Confusion matrix of expression analysis of thermal IR images with PCA.

	Ag	Ha	Fe	Ne	Sa
Ag	64.14	3	8.57	3	21.29
Ha	2.73	74.79	3.97	12.44	6.07
Fe	3.01	3.72	78.37	11.49	3.41
Ne	1.25	7.20	5.08	86.47	-
Sa	5.84	-	-	4.82	89.34
Avg	78.62				

Table 3.14: Confusion matrix of expression analysis of thermal IR images with PCA-EMC.

	Ag	Ha	Fe	Ne	Sa
Ag	92.28	-	4.29	-	3.43
Ha	-	83.82	3.03	9.26	3.89
Fe	-	3.86	90.58	5.56	-
Ne	-	9.90	1.36	88.74	-
Sa	3.76	6.29	-	7.80	82.15
Avg	87.51				

3.6.4 Evaluation of the Thermal Image Database

Before classifying images to emotions, the preprocessing is similar to the preprocessing for the visible images. The PCA, EMC, PCA-EMC also used to classify the emotions. In this experiment, we use 330 images of 22 subjects for 5 expressions. We used 40 percent images and 60 percent images in total images to test and train, respectively. From table 3.13, 3.14, 3.15, EMC method, suitable for thermal IR images, gave very high classify rate.

3.7 Conclusions

In this chapter, we have presented a KTFE database for the estimation emotion and analysis of spontaneous emotions. The database contains visible and thermal IR videos of 30 participants, where each participant watched the designed video, induced emotions and

Table 3.15: Confusion matrix of expression analysis of thermal IR images with EMC.

	Ag	Ha	Fe	Ne	Sa
Ag	83.08	-	-	-	16.92
Ha	-	99.36	0.06	0.5	0.08
Fe	-	-	100	-	-
Ne	-	-	-	100	-
Sa	6.54	-	-	0.07	93.39
Avg	95.17				

made self-reports about their true emotion. This new database offers several advantages with respect to the previously existing databases. Firstly, this is one of the first natural spontaneous visible and thermal IR videos. These databases will allow researchers on facial expressions and emotions to have more approaches more realistic; they can be used for visible, infrared, or multi-spectral natural emotion estimation. The database also contains the facial temperatures of subjects, thereby providing the potential for emotion estimation. Secondly, this database already fixed some mistakes which the former database met when they did experiment settings such as the time lag phenomenon; Thirdly, we did the analysis of the effectiveness of eliciting video and the results of analysis show our selected emotion-eliciting videos are suitable and effective to induce the emotions; Fourthly, we analyzed the temperature trend of each emotion. The results may be helpful for inference human emotion when we use thermal IR information. Finally, we also carried out an elementary assessment of the usability of the spontaneous sub-database, using three baseline algorithms for visible emotion estimation and thermal emotion estimation. The results on thermal IR data give us a promising future on facial research. In future, we will continue developing our data and working on thermal IR data to contribute better results.

Chapter 4

Estimation of Human Emotion by Reducing the Effect of Eyeglasses

In recent years, research on human emotion estimation using thermal IR image has appealed to many researchers due to its invariance to visible illumination changes. Although infrared image is superior to visible image in its invariance to illumination changes and appearance differences, it has difficulties in handling transparent glasses in the thermal IR spectrum. As a result, when using infrared image for the analysis of human facial information, the regions of eyeglasses are dark and eyes thermal IR information is not given. A single thermal IR frame can not give the exact human emotion, hence we suggest a method using thermal IR frame sequence. We propose a temperature space method to correct eyeglasses effect using the thermal IR facial frame sequence and then thermal-Principal Component Analysis (t-PCA), norm-Eigenspace method based on class feature (n-EMC) and PCA & EMC to classify human emotions from the corrected thermal IR images. We collected the Kotani Thermal Facial Emotion (KTFE) database and performed the experiments, which show the improved accuracy rate in estimating human emotions.

4.1 Introduction

Nowadays, Human Computer Interaction (HCI) is a very attractive research area in computer vision. One of the key researches in HCI is to detect the inner emotions through human faces by performing automatic analysis of facial expressions. Many previous works have proposed towards developing facial emotion estimation [96]. However, we still lack an accurate and robust facial emotion estimation method to be deployed in uncontrolled environments. Several factors that affect facial emotion estimation include pose variations, occlusions, and most importantly, illumination changes [96]. Therefore, it is a new and imaginative way to use IR image, which is not sensitive to light condition, to fill the gap in the human emotion estimation field. Besides, human emotions could be manifested by changing temperature of face skin which is obtained by IR camera. Consequently, thermal IR image gives us more information to help us robustly estimate the human emotions. Recently, a number of studies have demonstrated that thermal IR image offers a promising alternative to visible image in facial emotion estimation problems by better handling the visible illumination changes. Y.Yoshitomi et al. used two dimensional detection of temperature distribution on the face using infrared rays [33]. Based on studies in the field of psychology, several blocks on the face are chosen for measuring the local temperature difference. With Back Propagation Neutral Network, the facial expression is recognized. The recognition accuracy reaches 90% with neutral, happy, surprising and sad expressions. However, the testing database is obtained from only one female frontal view. Y. Yoshimomi generated feature vectors by using a two-dimensional Discrete Cosine Transformation (2D-DCT) to transform the grayscale values of each block in the facial area of an image into their frequency components, and used them to recognize five expressions, including angry, happy, neutral, sad, and surprise. The mean expression accuracy is 80% with four test subjects [90]. Y.Koda et al. used the idea from [90] and added a proposed method for efficiently updating of training data, by only updating the training data with happy and neutral facial expression after an interval [91]. The expression accuracy increased from 80% to 87% with this new approach. Sophie Jarlier et al. extracted the features as representative temperature maps of nice action unit

(AUs) and used K-nearest neighbor to classify seven expressions [89]. The database for testing has four persons and the accuracy rate is 56.4%. M.M.Khan et al. suggested using Facial Thermal Feature Points (FTFPs), which are defined as facial points that undergo significant thermal IR changes in presenting an expression, and used Linear Discriminant Analysis (LDA) to classify intentional facial expressions based on Thermal Intensity Values (TIVs) recorded at the Facial Thermal Feature Points (FTFPs) [23]. The database has 16 persons with 5 expressions and the accuracy rate ranges from 66.3% to 83.8%. L.Trujillo et al. proposed using a local and global automatic feature localization procedure to perform facial expression in thermal IR images. They used PCA to reduce the dimension and interest point clustering to estimate facial feature localization and Support Vector Machine (SVM) to classify three expressions [24]. B.Hernandez et al. used SVM to classify the expressions surprise, happy, neutral from two inputs. First input consists of selection of a set of suitable regions where the feature extraction is performed, second input is the Gray Level Co-occurrence Matrix used to compute region descriptors of the IR images [25]. B.R.Nhan et al. extracted time, frequency and time-frequency features from thermal IR data to classify the natural responses in terms of subject-indicated levels of arousal and valence stimulated by the International Affective Picture System [26].

All these studies with thermal IR image have shown that human emotion states are related to the facial skin temperature property. However, to our knowledge, most approaches that use the extracted features from a single thermal IR image may lose some useful information which could be contained in the sequences. Although there are many significant advantages when we use thermal IR image, IR have several drawbacks. One of drawbacks is eyeglasses. Glass is opaque to IR and the sensitivity of thermal IR to facial occlusions is decreased by eyeglasses. This is because objects made of glasses act as a temperature screen, completely occluding the parts located behind them. To eliminate the effect of presence of eyeglasses, we propose a temperature space method for thermal IR data. We will also reduce the impact of ambient temperature by normalizing it between frames. To estimate seven emotions, we use t-PCA, PCA & EMC, and n-EMC to extract feature vectors and then find the similarity between the testing data and training data.

4.2 Methods

In this section, we propose a method to reduce the effect of eyeglasses in thermal IR facial frame sequence and then apply three our proposed methods t-PCA, n-EMC, and PCA & EMC to analyze emotions. Before applying temperature space, to avoid the temperature change of ambience from frame to frame, we calculate the mean of ambient point temperatures of each frame. Then we find the difference of the mean of each frame to that of its previous frame and update the temperature of all points of each frame by subtracting those temperatures with the difference.

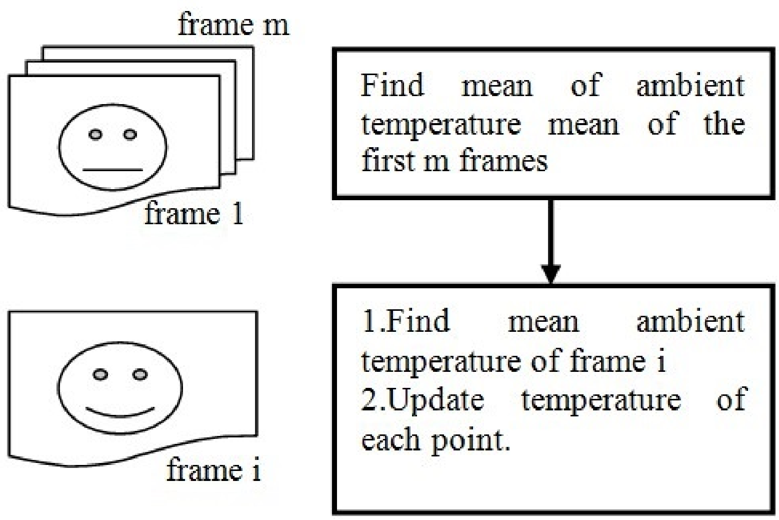


Figure 4.1: Reducing the effect of changing ambient temperature

4.2.1 Temperature Space

We propose using temperature space to support the temperature analysis of human face area. Based on temperature data and observation, we classify images into two main spaces, i.e. non-face space and face space. Non-face space is for the background area, which further includes two sub-spaces. In face space, there are three sub-spaces because the eyebrows and nose are usually cold whereas the cheek is warm [10], and the forehead is usually the warmest. Using the spaces, we find eyeglasses area and replace each point of them by the mean of image temperature.

Let f and g be maps from image($I \subset R^2$) to temperature space ($T \subset R$) and image ($I \subset R^2$) to glass space ($G \subset R^2$) respectively, as the following:

$$f : I \to T$$
$$(i, j) \mapsto f(i, j)$$

$$g : T \to G$$
$$(i, j) \mapsto g \circ f(i, j)$$

$$g \circ f(i,j) = \begin{cases} 0, & \text{if } (i,j) \in glass. \\ f(i,j), & \text{if } (i,j) \notin glass. \end{cases} \tag{4.1}$$

where I, T, G are image space, temperature space and glass space, respectively.
We calculate the temperature space for image and face as following:

$$\Delta T_I = T^I_{Max} - T^I_{Min}; \delta T_I = \Delta T_I / h \tag{4.2}$$

where T^I_{Max}, T^I_{Min} are maximum and minimum of temperature of each image, respectively.

$$L_k^I = \{(i,j) \in I | T_{Min}^I + \delta T_I * (k-1) \leq f(i,j) < T_{Max}^I - \delta T_I * (h-k)\} \tag{4.3}$$

where $k \in (\overline{1,h})$,

$$\Delta T_F = T_{Max}^F - T_{Min}^F; \delta T_F = \Delta T_F / h \tag{4.4}$$

where T_{Max}^F, T_{Min}^F are maximum and minimum of temperature of each human face area, respectively.

$$L_l^F = \{(i,j) \in F | T_{Min}^F + \delta T_F * (l-1) \leq f(i,j) < T_{Max}^F - \delta T_F * (h-l)\} \tag{4.5}$$

where $l \in (\overline{1,h})$, $h = 5$;

$$L_0^F = \{(i,j) \in F | f(i,j) < T_{Min}^F\} \tag{4.6}$$

Based on (1.3), (1.5), (1.6), we find the glass area as following:

$$glass = \{(i,j) | f(i,j) \in [sup\{inf L_2^I, inf L_0^F\}, inf\{L_3^I, inf L_1^F\}]\} \tag{4.7}$$

After using our proposed method to find eyeglasses area in thermal IR data, to reduce the effect of eyeglasses, we replace the temperature of the areas by the mean of ambient temperature.

4.2.2 Estimate of Human Emotion

To estimate the true emotion, using single frame can not give the exact emotion, hence we use the sequence of thermal IR frames. Therefore, after reducing the effect of eyeglasses, to apply for sequence of thermal IR frames, we calculate the accumulate of the discriminant of frame m and frame m + idx.

Let call $F^* = F \backslash G \subset R^2$

$$f^* : F^* \to T$$
$$(i,j) \mapsto f^*(i,j)$$

We calculate $\sum_m \parallel f_m^*(i,j) - f_{m+idx}^*(i,j) \parallel, \forall (i,j) \in F^*$

To analyze the human emotion using thermal IR data of emotion, we propose three methods, thermal-Principal Component Analysis (t-PCA), norm-Eigenspace method based on class feature (n-EMC) and PCA & EMC.

To illustrate the feasibility of using eigen-space to fulfill facial emotion estimation task, thermal PCA (t-PCA) is modified from the PCA reconstruction method and evaluated over thermal IR data [113]. With PCA, the aim is to build a face space, including the basis vectors called principal components, which better describe the face images. PCA has several advantages over other face recognition schemes in its speed and simplicity [113]. We modified PCA to estimate facial emotion from thermal IR data.

Let F be a set of classes to be analyzed. Here, F is a set of all emotion classes. Assume that M_m^f thermal IR frames of training data are given as the facial temperature pattern for each class $f \in F$ where $F = \{anger, disgust, fear, happiness, neutral, sadness, surprise\}$. Let Γ_i^f be the $i-th$ facial temperature pattern where $i = \overline{1, M_f}$; the dimension of Γ_i^f, $n \times m$, is equal to the number of pixels in a thermal IR frame, and each element of Γ_i^f indicates the temperature of each pixel.

Compute the mean of training data $\Psi^f = \frac{1}{M^f}\sum_{i=1}^{M_f}\Gamma_i^f$ and let the normalized vector be $\Phi_i^f = \Gamma_i^f - \Psi^f$. We seek a set of M orthonormal vectors, u_k^f, which best describes the distribution of the training data. The kth vector, u_k^f is chosen by

$$\lambda_k^f = \frac{1}{M_f}\sum_{i=1}^{M_f}((u_k^f)^\tau \Phi_i^f)^2. \tag{4.8}$$

is a maximum, subject to $(u_l^f)^\tau u_k^f = \begin{cases} 1, & \text{if } l = k. \\ 0, & \text{if } otherwise. \end{cases}$

The eigenvectors and eigenvalues are the vector u_k^f and scalar λ_k^f of covariance matrix $C^f = \frac{1}{M_f}\sum_{i=1}^{M_f}(\Phi_i^f(\Phi_i^f)^\tau = A^f(A^f)^\tau$ where $A^f = [\Phi_1^f, \Phi_2^f, ..., \Phi_{M_f}^f]$

After obtaining eigenfaces from training data of each emotion, we map facial thermal IR training data to feature spaces by $\omega_i^f(train) = (u_i^f)^\tau(\Gamma^f - \Psi^f), i = \overline{1, \rho^f}$.

We use the idea that if the input frame is much similar to some emotion training set, the reconstructed data will has less distortion than the data reconstructed from other eigenvectors of training emotions [113]. For each testing facial thermal IR frame Γ_{test}, firstly we project it onto the eigenfaces of each class.

$\omega_{test}^f = (U^f)^\tau(\Gamma_{test} - \Psi^f)$, where $U^f = (u_i^f), i = \overline{1, \rho^f}$.

Secondly, for each emotion, we find the feature which is most similar to the testing projected vector by calculate the angle between vector of training feature space and testing projected vector.

$$\beta^f = \operatorname*{argmax}_i \frac{\omega_{test}^f \omega_i^f(train)}{\|\omega_{test}^f\|\|\omega_i^f(train)\|}; i = \overline{1, \rho^f}. \tag{4.9}$$

Thirdly, we find reconstruction of the testing data by the obtained feature in each class.

$\Gamma_{reconst}^f = U^f \omega_{\beta^f}^f + \Psi^f$

Finally, we choose an emotion of which reconstruction of the testing data is the most similarity to testing data.

$$\gamma = \operatorname*{argmax}_f \frac{\Gamma_{test}\Gamma_{reconst}^f}{\|\Gamma_{test}\|\|\Gamma_{reconst}^f\|}; f = \overline{1, 7} \tag{4.10}$$

The second valuation to estimate human emotion uses n-EMC over thermal IR data. n-EMC is modified from EMC [114]. The difference between EMC and n-EMC is formulation to calculate the difference between the within-class and between-class variance.

In mathematics, with n-EMC, instead of finding the eigenvectors, u_k^f and eigenvalues λ_k^f of covariance matrix $C^f = \frac{1}{M_f}\sum_{i=1}^{M_f}(\Phi_i^f(\Phi_i^f)^\tau = A^f(A^f)^\tau$ where $A^f = [\Phi_1^f, \Phi_2^f, ..., \Phi_{M_f}^f]$, we find eigenvectors, u_k^f and eigenvalues λ_k^f of matrix $S = \|S_B - S_W\|_2$ where

$$M = \sum_{f \in F} M_f \tag{4.11}$$

$$\Psi^f = \frac{1}{M^f}\sum_{i=1}^{M_f}\Gamma_i^f; \Psi = \frac{1}{M}\sum_{f\in F}\sum_{i=1}^{M_f}\Gamma_i^f \tag{4.12}$$

$$S_B = \frac{1}{M}\sum_{f\in F} M_f\|\Psi_f - \Psi\|_2\|\Psi_f - \Psi\|_2^\tau. \tag{4.13}$$

$$S_W = \frac{1}{M}\sum_{f\in F}\sum_{i=1}^{M_f}\|\Gamma_i^f - \Psi_f\|_2\|\Gamma_i^f - \Psi_f\|_2^\tau. \tag{4.14}$$

For each testing facial thermal IR frame Γ_{test}, firstly we project it onto the eigenfaces of each class.

$\omega_{test}^f = (U^f)^\tau(\Gamma_{test} - \Psi^f)$, where $U^f = (u_i^f), i = \overline{1,\rho^f}$.

Secondly, for each emotion, we find the feature which is most similar to the testing projected vector by calculate the angle between vector of training feature space and testing projected vector.

$$\beta^f = \underset{i}{\operatorname{argmax}} \frac{\omega_{test}^f\omega_i^f(train)}{\|\omega_{test}^f\|\|\omega_i^f(train)\|}; i = \overline{1,\rho^f}. \tag{4.15}$$

Finally, we choose an emotion which has maximum of β^f

$$\gamma = \underset{f}{\operatorname{argmax}}\,\beta^f; f = \overline{1,7} \tag{4.16}$$

With PCA & EMC, we compute the mean of training data $\Psi^f = \frac{1}{M^f}\sum_{i=1}^{M_f}\Gamma_i^f$ and let the normalized vector be $\Phi_i^f = \Gamma_i^f - \Psi^f$. We seek a set of M orthonormal vectors, u_k^f, which best describes the distribution of the training data. The kth vector, u_k^f is chosen by

$$\lambda_k^f = \frac{1}{M_f}\sum_{i=1}^{M_f}((u_k^f)^\tau\Phi_i^f)^2. \tag{4.17}$$

is a maximum, subject to $(u_l^f)^\tau u_k^f = \begin{cases} 1, & \text{if } l = k. \\ 0, & \text{if } otherwise. \end{cases}$

The eigenvectors and eigenvalues are the vector u_k^f and scalar λ_k^f of covariance matrix $C^f = \frac{1}{M_f}\sum_{i=1}^{M_f}(\Phi_i^f(\Phi_i^f)^\tau = A^f(A^f)^\tau$ where $A^f = [\Phi_1^f, \Phi_2^f, ..., \Phi_{M_f}^f]$

After obtaining eigenfaces from training data of each emotion, we map facial thermal IR training data to feature spaces by $\omega_i^f(train) = (u_i^f)^\tau(\Gamma^f - \Psi^f), i = \overline{1,\rho^f}$. Then, we apply EMC to obtained data and do the same previous n-EMC analysis.

4.3 Experiments

For database, we use KTFE (Kotani Thermal Facial Emotion) database [112]. This database contains 30 subjects who are Vietnamese, Japanese, Thai from 11 year-olds to 32 year-olds with seven emotions. In our experiments, we use only sequence of thermal

IR data to estimate human emotions. From 186 GB thermal data, we extracted 28 GB of sequence thermal IR data for seven emotions.

In our experiments, we use PCA, EMC, t-PCA, n-EMC, PCA & EMC with eyeglasses and with reducing the effect of eyeglasses. From the obtained database, we separate the training and testing data as 60% and 40 % of the total data.

Table 4.1 and table 4.2 shows the results of PCA with eyeglasses and with reducing the effect of eyeglasses. The average of accuracy of PCA with reducing the effect of eyeglasses for sequence of thermal IR images increases 5.42% in comparison with having eyeglasses. Especially, for the anger emotion, the average of accuracy increase 8.75 % with reducing the effect eyeglasses.

Table 4.3 and table 4.4 shows the results of t-PCA with eyeglasses and with reducing the effect of eyeglasses. The average of accuracy of t-PCA with reducing the effect of eyeglasses for sequence of thermal IR images increases 3.87% in comparison with having eyeglasses. Especially, for the neutral emotion, the average of accuracy increase 5.77 % with reducing the effect eyeglasses.

Table 4.5 and table 4.6 shows the results of EMC with eyeglasses and with reducing the effect of eyeglasses. The average of accuracy of EMC with reducing the effect of eyeglasses for sequence of thermal IR images increases 3.08% in comparison with having eyeglasses. Especially, for the anger emotion, the average of accuracy increase 4.19 % with reducing the effect eyeglasses.

Table 4.7 and table 4.8 shows the results of n-EMC with eyeglasses and with reducing the effect of eyeglasses. The average of accuracy of n-EMC with reducing the effect of eyeglasses for sequence of thermal IR images increases 2.51% in comparison with having eyeglasses. Especially, for the sadness emotion, the average of accuracy increase 2.96 % with reducing the effect eyeglasses.

Table 4.9 and table 4.10 shows the results of PCA & EMC with eyeglasses and with reducing the effect of eyeglasses.The average of accuracy of PCA & EMC with reducing the effect of eyeglasses for sequence of thermal IR images increases 4.13% in comparison with having eyeglasses. Especially, for the disgust emotion, the average of accuracy increase 11.09 % with reducing the effect eyeglasses.

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	75.63	-	-	11.88	5.00	-	7.50
Disgust	-	75.07	3.85	1.04	-	3.33	16.71
Fear	-	1.39	77.54	8.77	4.76	4.76	2.78
Happiness	-	-	-	78.41	-	21.59	-
Neutral	11.22	-	0.38	7.55	76.58	4.28	-
Sadness	1.26	-	4.47	6.55	6.91	80.81	-
Surprise	9.62	11.40	-	1.00	-	-	77.98
Average							**77.43**

Table 4.1: The confusion matrix of PCA with eyeglasses

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	84.38	-	-	9.38	-	-	6.25
Disgust	-	82.66	4.53	-	-	-	12.81
Fear	3.16	0.01	83.47	5.83	2.41	5.19	0.01
Happiness	-	0.11	2.00	81.50	1.00	15.39	-
Neutral	7.25	-	2.34	5.17	82.29	2.94	-
Sadness	-	0.33	4.67	9.58	0.11	85.31	-
Surprise	2.98	15.57	-	1.12	-	-	80.32
Average							**82.85**

Table 4.2: The confusion matrix of PCA with reducing the effect of eyeglasses

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	79.55	6.82	-	2.27	4.55	6.82	-
Disgust	-	79.69	-	1.56	-	-	18.75
Fear	-	1.04	81.37	-	7.98	7.52	2.08
Happiness	-	-	-	82.95	-	17.05	-
Neutral	6.11	-	2.86	4.47	79.20	7.37	-
Sadness	1.53	-	3.37	3.85	6.38	84.87	-
Surprise	5.67	15.76	-	-	-	-	78.56
Average							**80.88**

Table 4.3: The confusion matrix of t-PCA with eyeglasses

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	84.67	2.14	2.90	3.99	-	2.14	4.16
Disgust	-	84.24	4.12	-	-	-	11.64
Fear	-	0.50	84.35	6.61	3.83	3.71	1.00
Happiness	-	-	-	86.36	-	13.64	-
Neutral	0.84	-	1.51	1.15	84.97	11.54	-
Sadness	-	0.56	2.22	9.03	0.18	88.02	-
Surprise	3.72	14.88	-	0.75	-	-	80.65
Average							**84.75**

Table 4.4: The confusion matrix of t-PCA with reducing the effect of eyeglasses

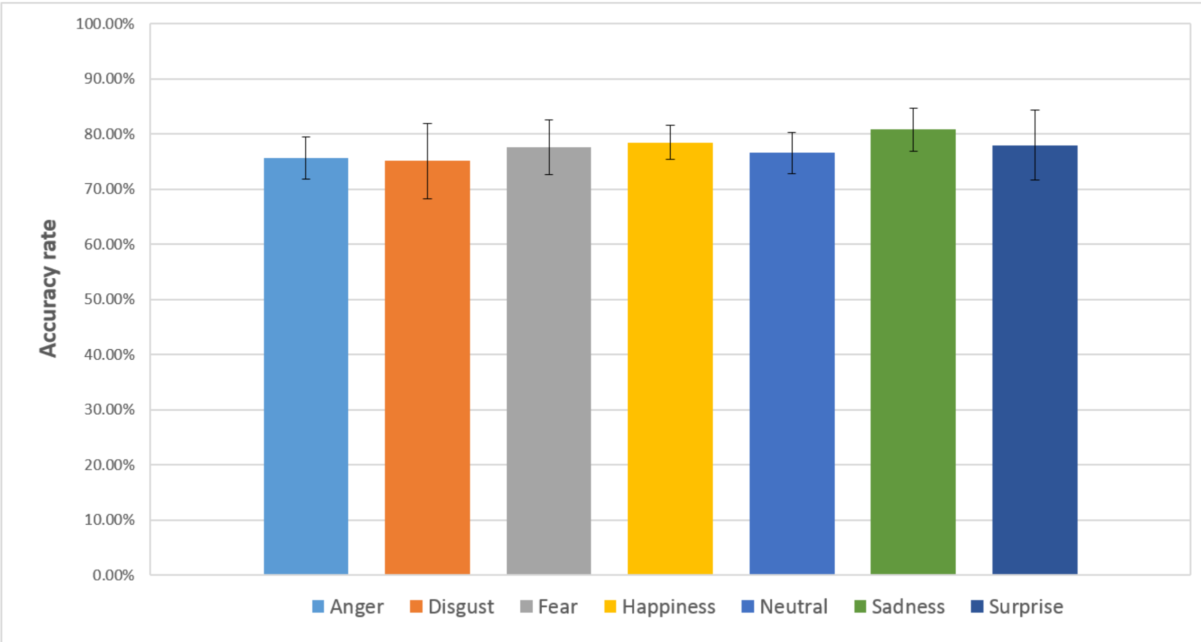


Figure 4.2: The emotion estimation results of PCA with eyeglasses

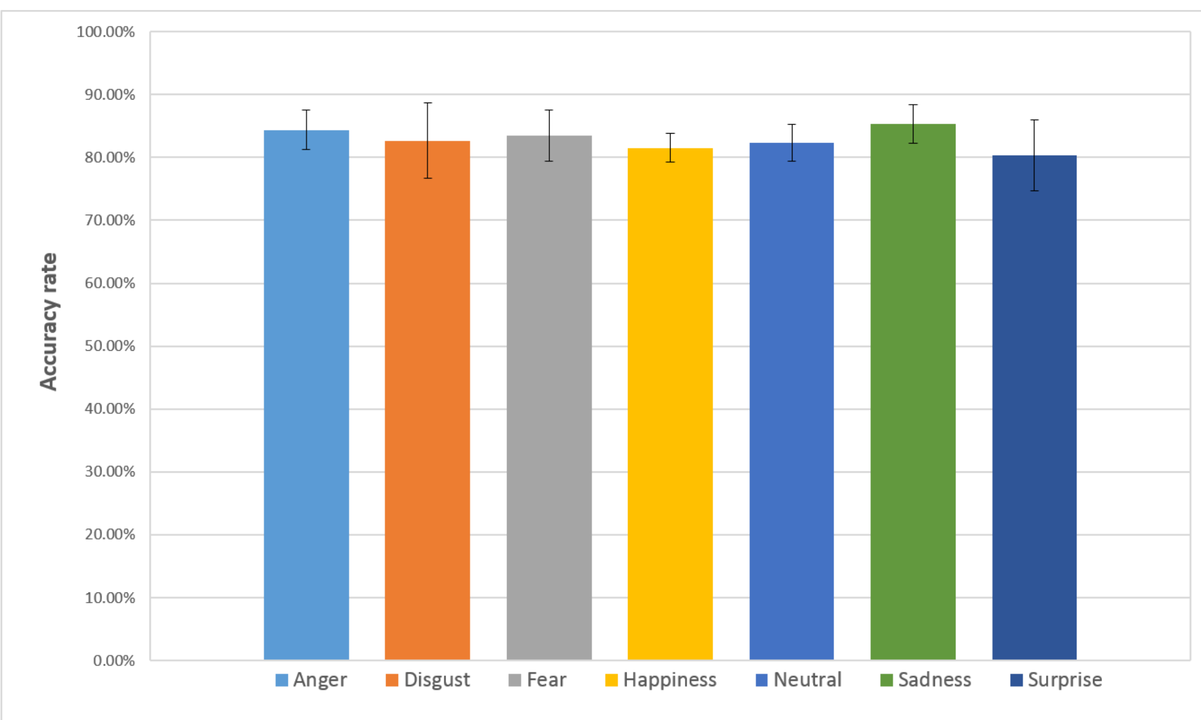


Figure 4.3: The emotion estimation results of PCA with reducing the effect of eyeglasses

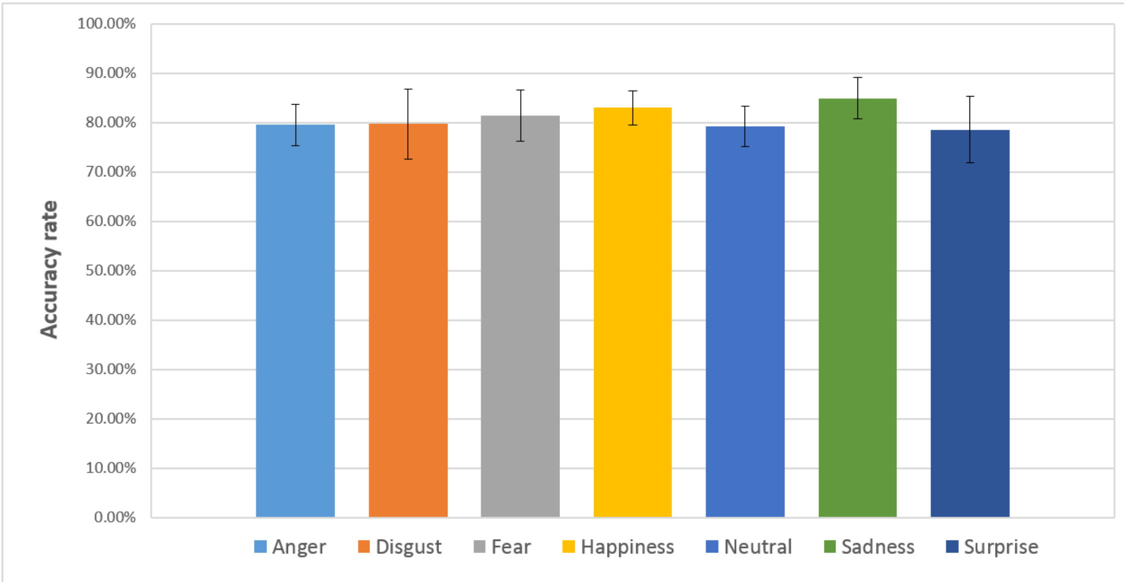


Figure 4.4: The emotion estimation results of t-PCA with eyeglasses

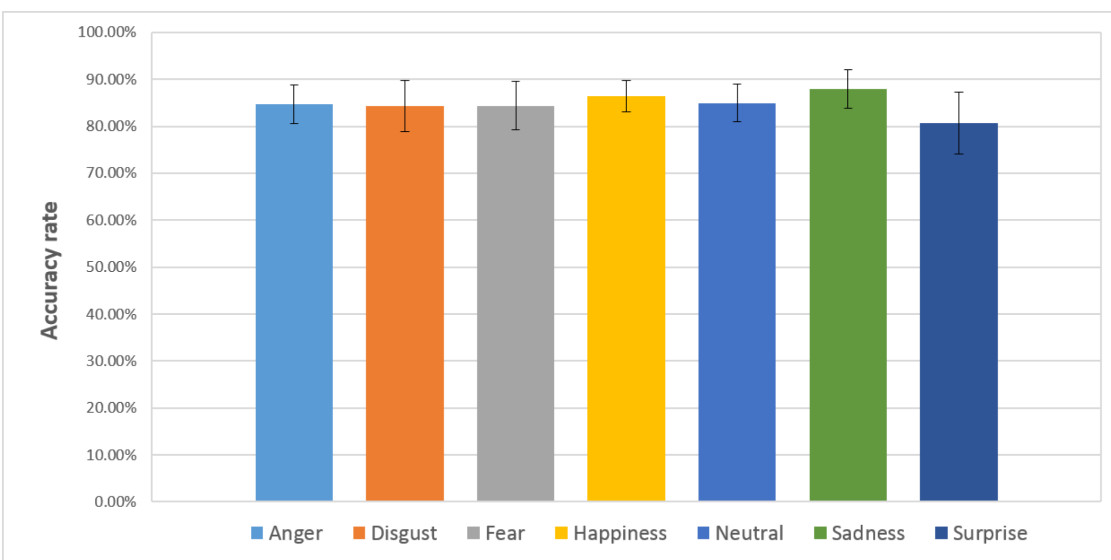


Figure 4.5: The emotion estimation results of t-PCA with reducing the effect of eyeglasses

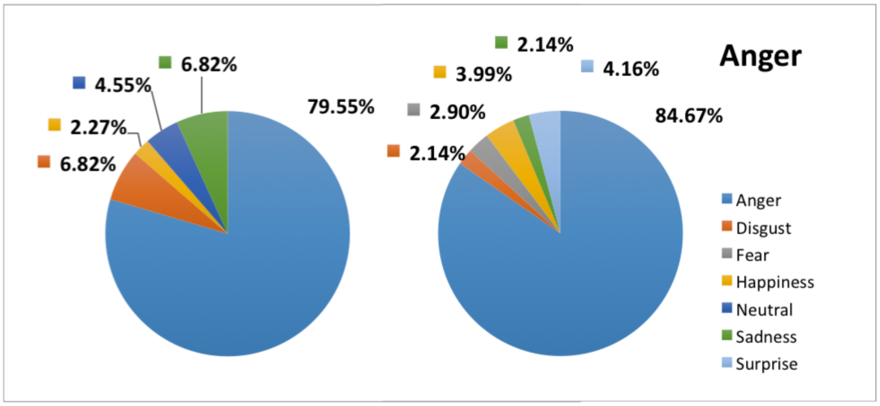


Figure 4.6: The comparison of anger estimation of t-PCA between with eyeglasses and with reducing the effect of eyeglasses

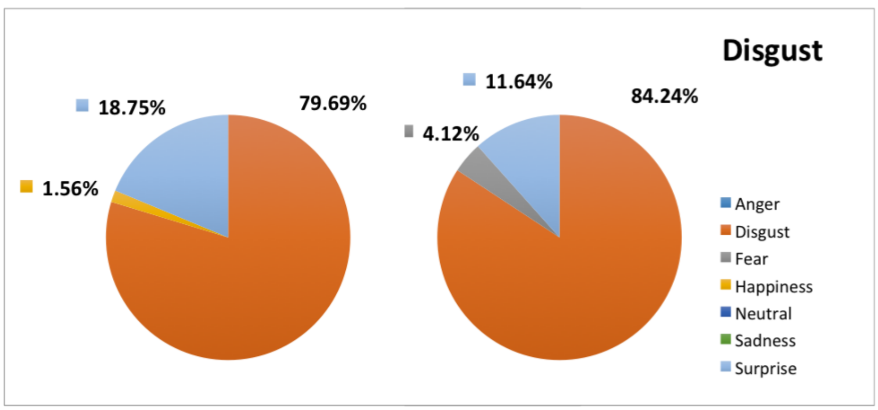


Figure 4.7: The comparison of disgust estimation of t-PCA between with eyeglasses and with reducing the effect of eyeglasses

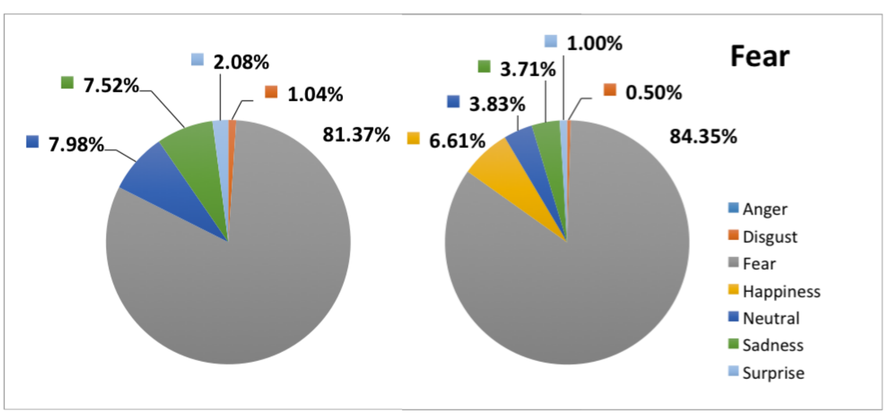


Figure 4.8: The comparison of fear estimation of t-PCA between with eyeglasses and with reducing the effect of eyeglasses

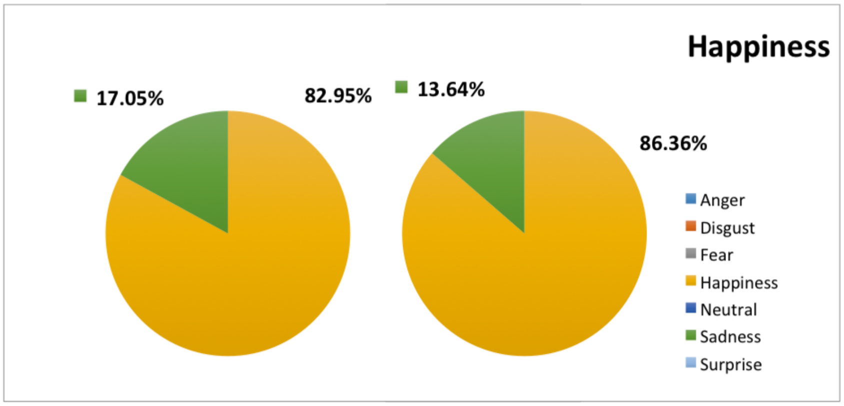


Figure 4.9: The comparison of happiness estimation of t-PCA between with eyeglasses and with reducing the effect of eyeglasses

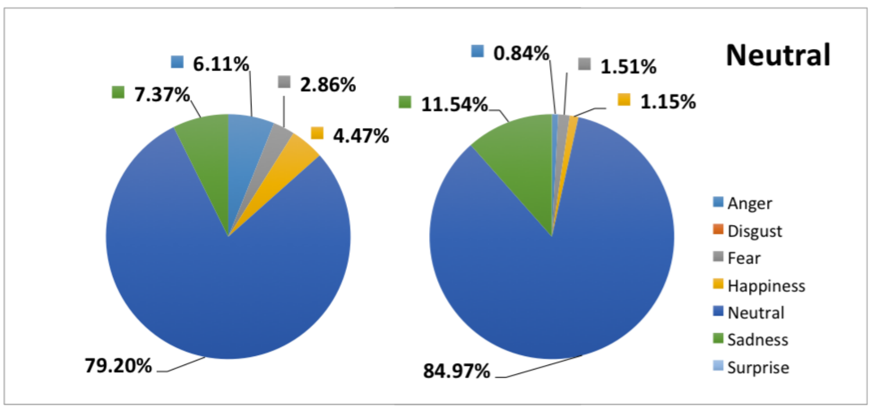


Figure 4.10: The comparison of neutral estimation of t-PCA between with eyeglasses and with reducing the effect of eyeglasses

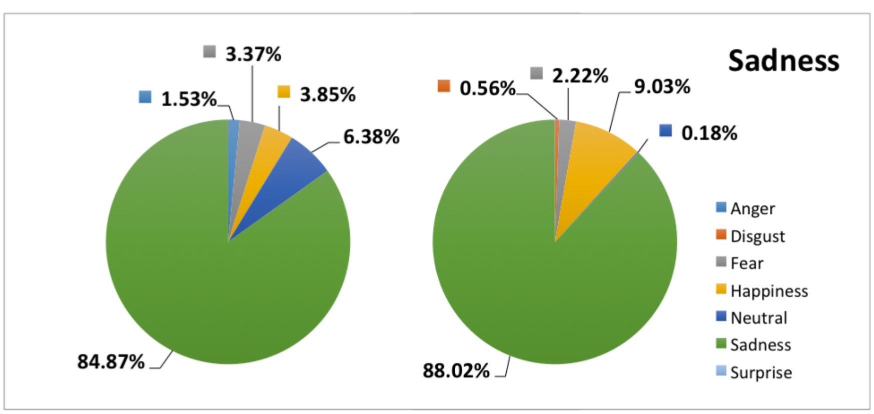


Figure 4.11: The comparison of sadness estimation of t-PCA between with eyeglasses and with reducing the effect of eyeglasses

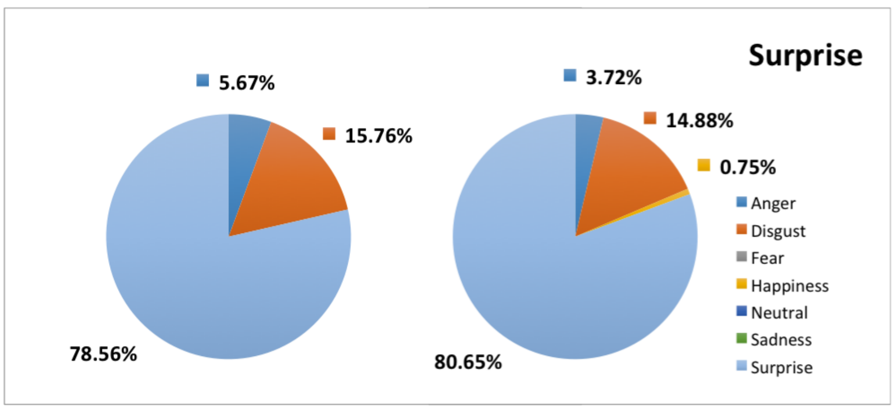


Figure 4.12: The comparison of surprise estimation of t-PCA between with eyeglasses and with reducing the effect of eyeglasses

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	80.98	3.58	-	8.51	-	6.93	-
Disgust	-	78.92	7.50	-	-	-	13.58
Fear	3.76	-	79.21	8.31	0.67	8.05	-
Happiness	-	-	-	84.32	-	15.68	-
Neutral	-	-	10.21	-	81.75	8.05	-
Sadness	-	-	16.45	5.05	-	78.49	-
Surprise	2.30	4.30	1.15	0.01	0.01	15.98	76.32
Average							**80.00**

Table 4.5: The confusion matrix of EMC with eyeglasses

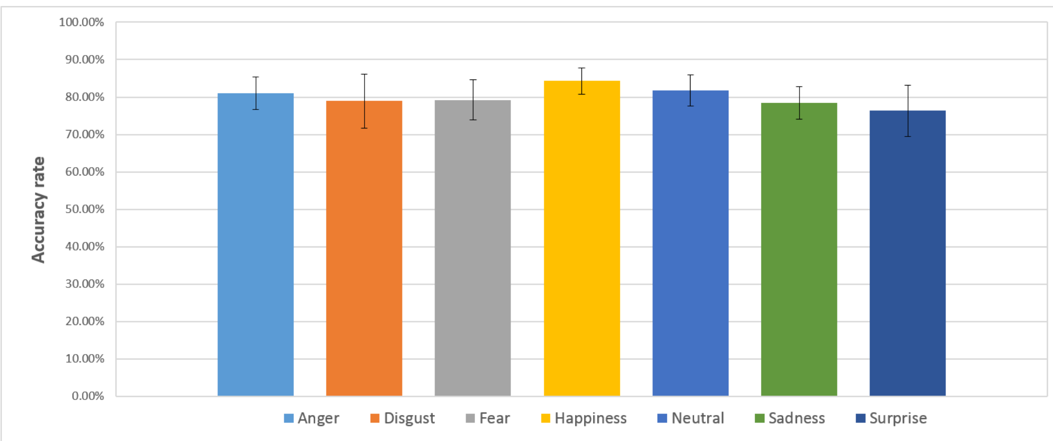


Figure 4.13: The emotion estimation results of EMC with eyeglasses

4.4 Conclusions

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	85.17	6.48	-	6.48	-	1.88	-
Disgust	3.37	81.67	-	-	-	-	14.96
Fear	3.68	-	80.97	5.83	1.02	8.52	-
Happiness	-	-	-	87.43	-	12.57	-
Neutral	-	-	6.97	-	84.33	8.71	-
Sadness	-	-	13.02	4.42	-	82.56	-
Surprise	0.96	3.57	0.48	-	-	15.58	79.42
Average							**83.08**

Table 4.6: The confusion matrix of EMC with reducing the effect of eyeglasses

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	82.33	2.28	-	7.27	-	8.13	-
Disgust	-	79.73	9.17	-	-	-	11.10
Fear	4.85	-	79.96	3.75	0.90	10.53	-
Happiness	-	-	-	85.99	-	14.01	-
Neutral	-	-	7.50	-	82.69	9.81	-
Sadness	-	-	17.16	3.16	-	79.68	-
Surprise	2.68	5.72	1.34	0.01	0.01	12.48	77.83
Average							**81.17**

Table 4.7: The confusion matrix of n-EMC with eyeglasses

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	85.41	6.55	-	6.54	-	1.50	-
Disgust	2.34	82.21	-	-	-	-	15.45
Fear	3.56	-	81.10	5.83	1.13	8.39	-
Happiness	-	-	-	88.61	-	11.39	-
Neutral	12.32	-	-	-	85.54	2.14	-
Sadness	-	-	13.82	3.54	-	82.64	-
Surprise	1.90	7.14	0.95	-	-	9.74	80.26
Average							**83.68**

Table 4.8: The confusion matrix of n-EMC with reducing the effect of eyeglasses

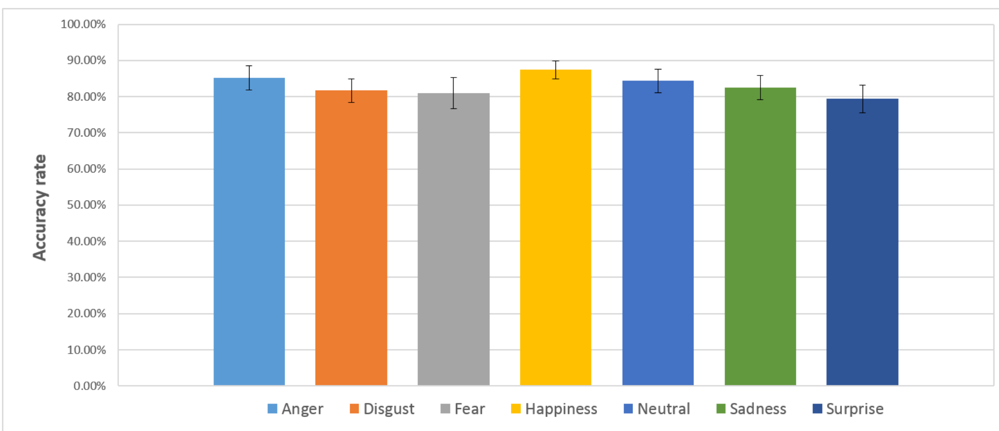


Figure 4.14: The emotion estimation results of EMC with reducing the effect of eyeglasses

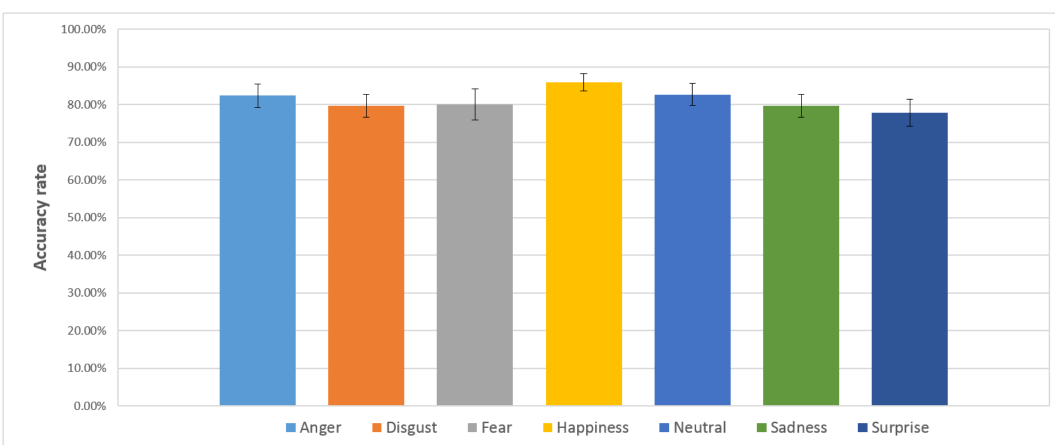


Figure 4.15: The emotion estimation results of n-EMC with eyeglasses

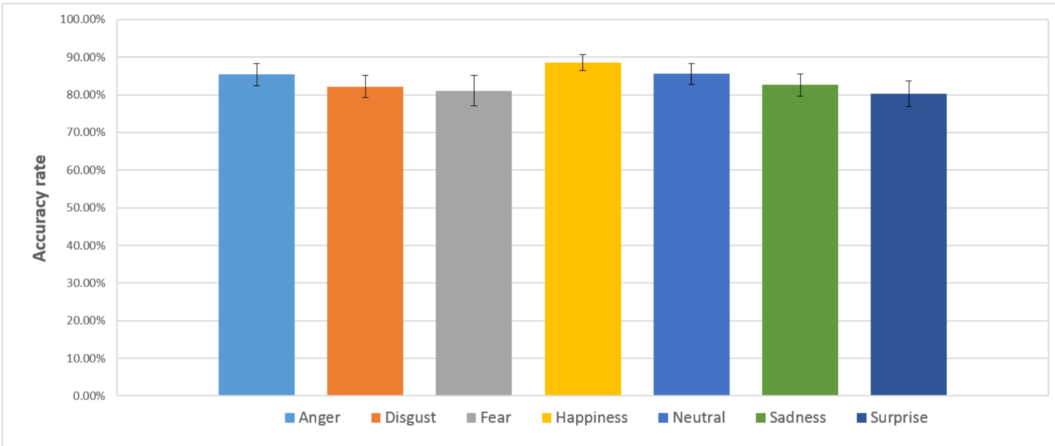


Figure 4.16: The emotion estimation results of n-EMC with reducing the effect of eyeglasses

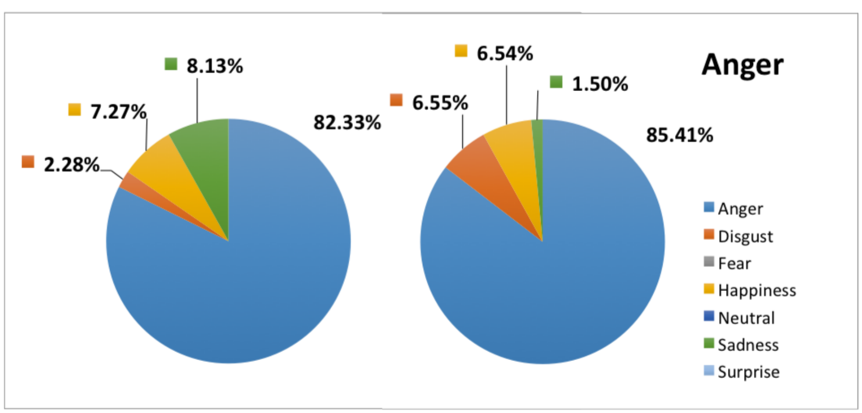


Figure 4.17: The comparison of anger estimation of n-EMC between with eyeglasses and with reducing the effect of eyeglasses

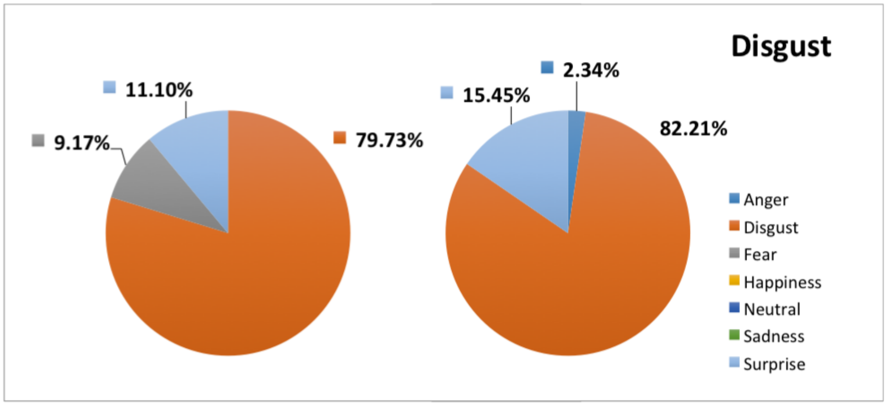


Figure 4.18: The comparison of disgust estimation of n-EMC between with eyeglasses and with reducing the effect of eyeglasses

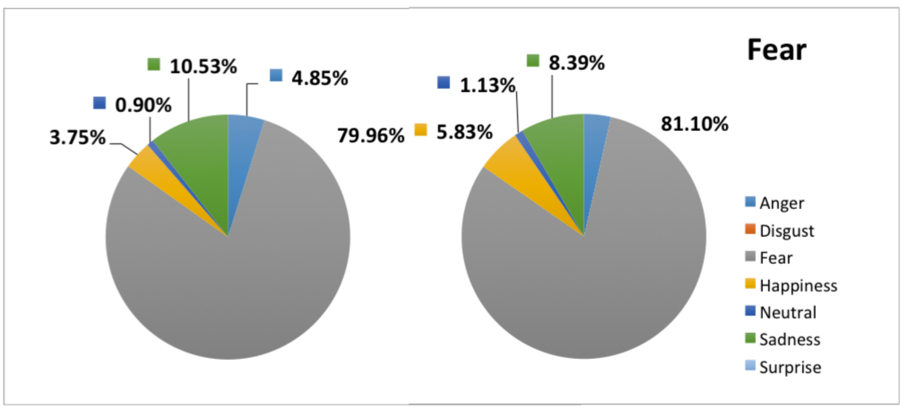


Figure 4.19: The comparison of fear estimation of n-EMC between with eyeglasses and with reducing the effect of eyeglasses

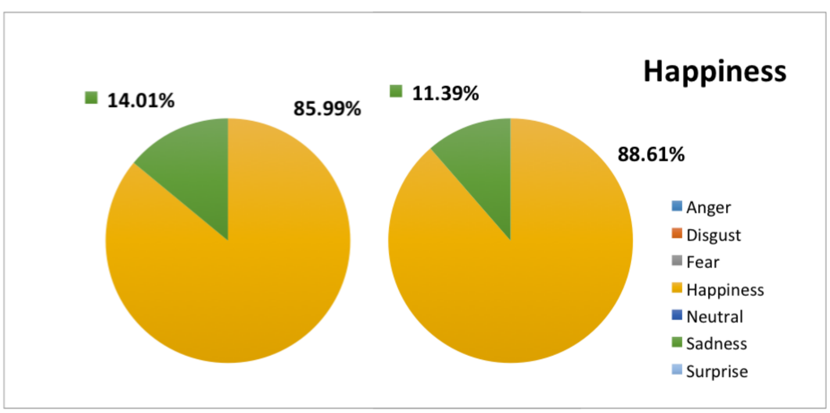


Figure 4.20: The comparison of happiness estimation of n-EMC between with eyeglasses and with reducing the effect of eyeglasses

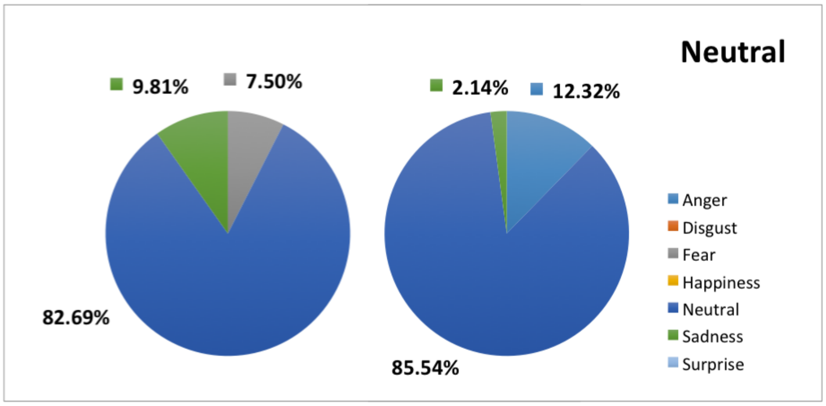


Figure 4.21: The comparison of neutral estimation of n-EMC between with eyeglasses and with reducing the effect of eyeglasses

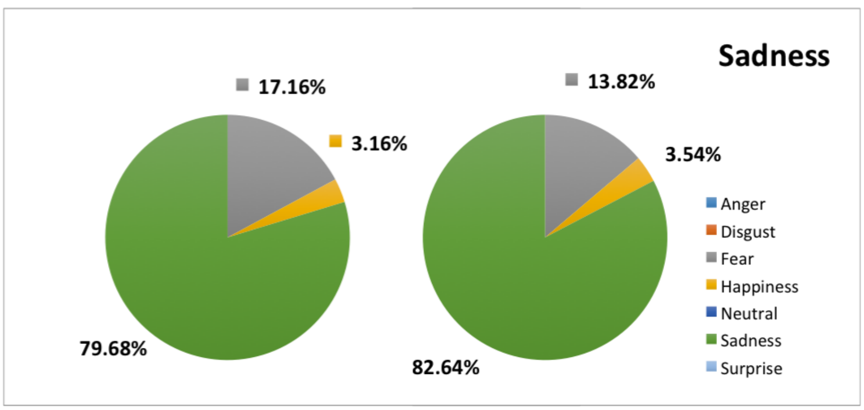


Figure 4.22: The comparison of sadness estimation of n-EMC between with eyeglasses and with reducing the effect of eyeglasses

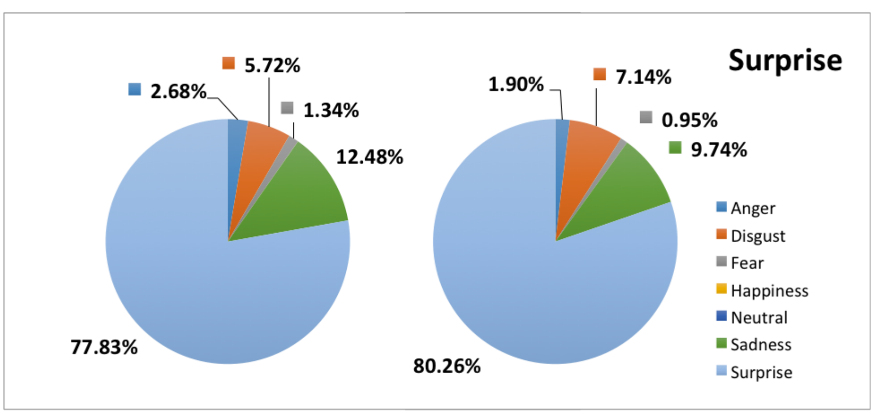


Figure 4.23: The comparison of surprise estimation of n-EMC between with eyeglasses and with reducing the effect of eyeglasses

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	62.87	7.86	9.01	4.69	-	12.77	2.80
Disgust	-	58.62	9.34	17.72	-	8.22	6.10
Fear	5.56	-	60.10	15.38	4.80	10.56	3.60
Happiness	-	0.82	7.11	65.65	5.21	18.51	2.70
Neutral	9.04	-	13.48	12.12	62.73	2.63	-
Sadness	10.84	-	11.31	12.16	2.69	63.00	-
Surprise	6.56	34.33	-	0.50	-	-	58.61
Average							**61.65**

Table 4.9: The confusion matrix of PCA & EMC with eyeglasses

This chapter describes a novel way to reduce the effect of changing ambient temperature which occurs during the data acquisition and the proposed method to reduce the effect of eyeglasses using temperature space in thermal IR data sequence. To estimate human emotion and prove the efficiency of the proposed method, three our proposed methods, namely t-PCA, n-EMC and PCA& EMC are used with KTFE database. Since the eyeglasses areas are replaced with the averaged-ambient temperature, these is no more effect of ambient temperature to these areas. The experiment with after removing glasses and before removing glasses shows the increasing of accuracy rate. Specially, with t-PCA, the accuracy of PCA & EMC emotion increases 11.9%. In general, the accuracy of each emotion and each method has been improved with the correction of eyeglass areas.

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	63.90	8.36	7.46	6.36	-	11.74	2.18
Disgust	-	69.70	5.27	4.55	-	-	20.48
Fear	-	-	64.85	16.72	5.34	5.45	7.64
Happiness	1.20	-	4.41	69.49	5.78	19.12	-
Neutral	8.46	-	11.58	11.60	66.09	2.27	-
Sadness	3.56	1.50	7.13	13.26	9.06	65.49	-
Surprise	5.74	32.57	-	0.72	-	-	60.97
Average							**65.78**

Table 4.10: The confusion matrix of PCA & EMC with reducing the effect of eyeglasses

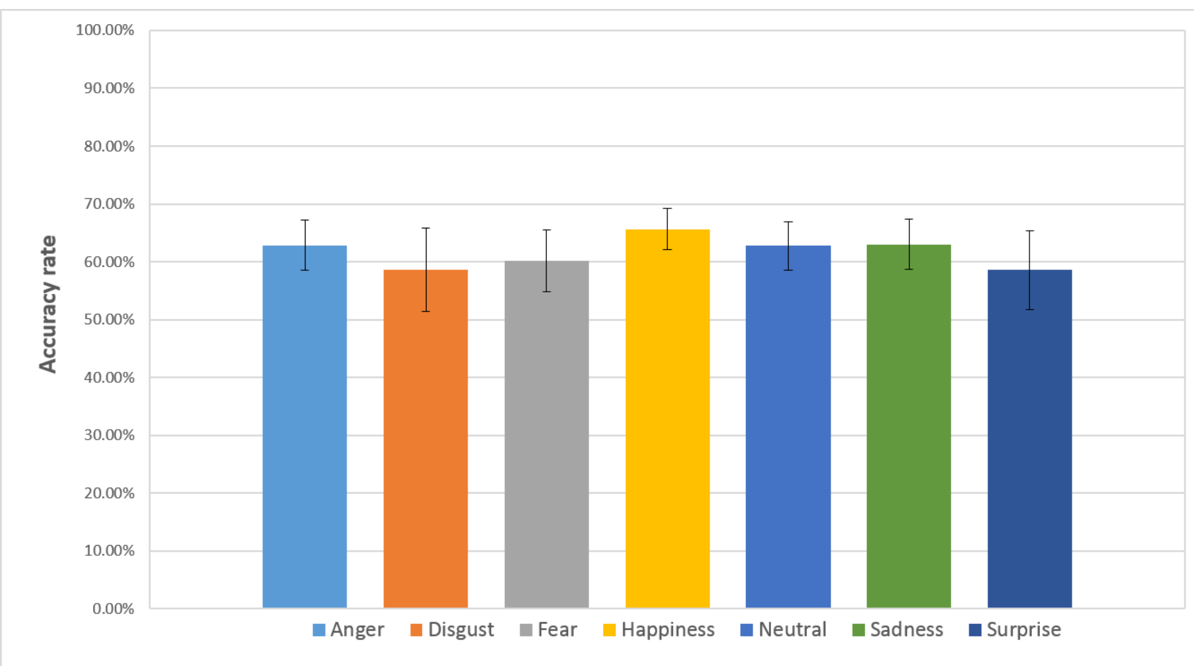


Figure 4.24: The emotion estimation results of PCA & EMC with eyeglasses

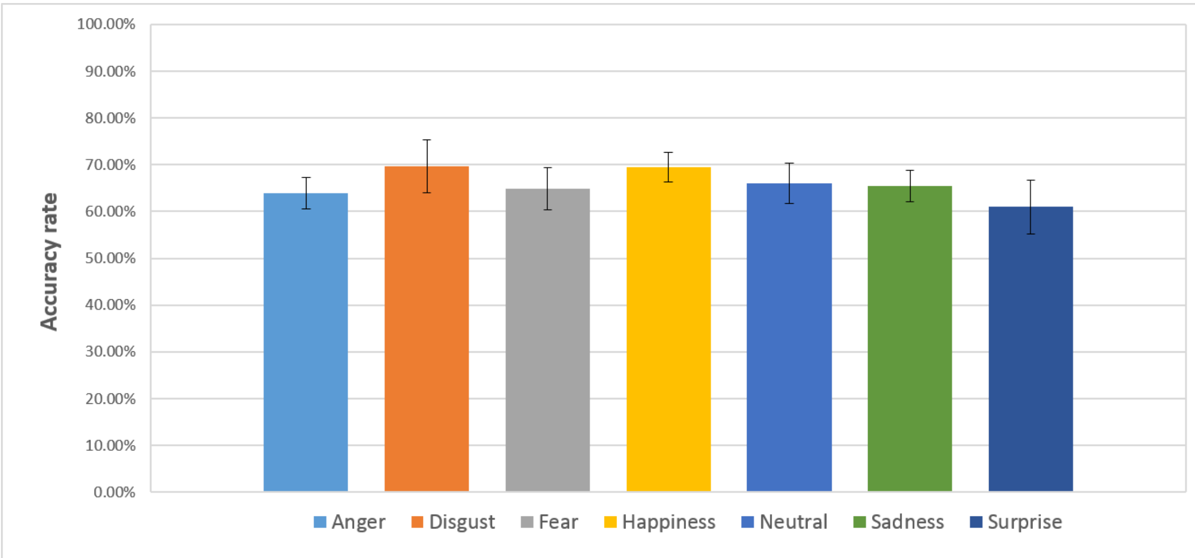


Figure 4.25: The emotion estimation results of PCA & EMC with reducing the effect of eyeglasses

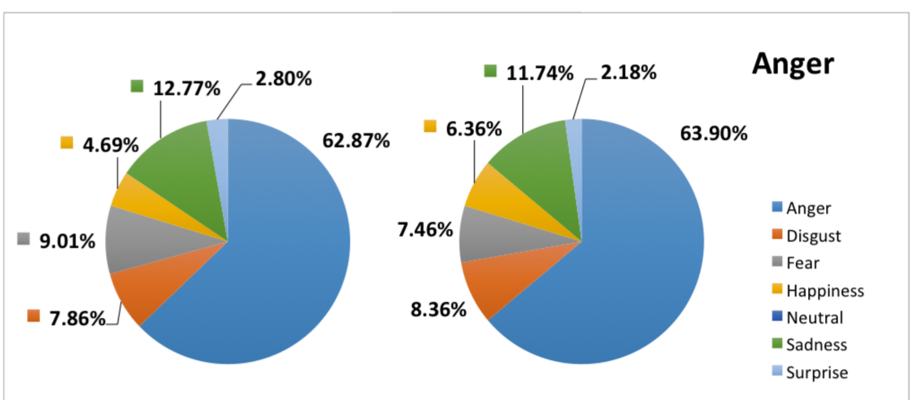


Figure 4.26: The comparison of anger estimation of PCA & EMC between with eyeglasses and with reducing the effect of eyeglasses

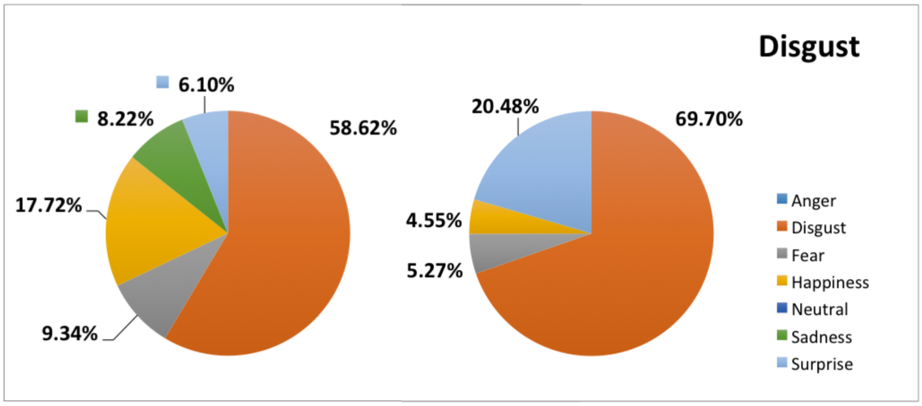


Figure 4.27: The comparison of disgust estimation of PCA & EMC between with eyeglasses and with reducing the effect of eyeglasses

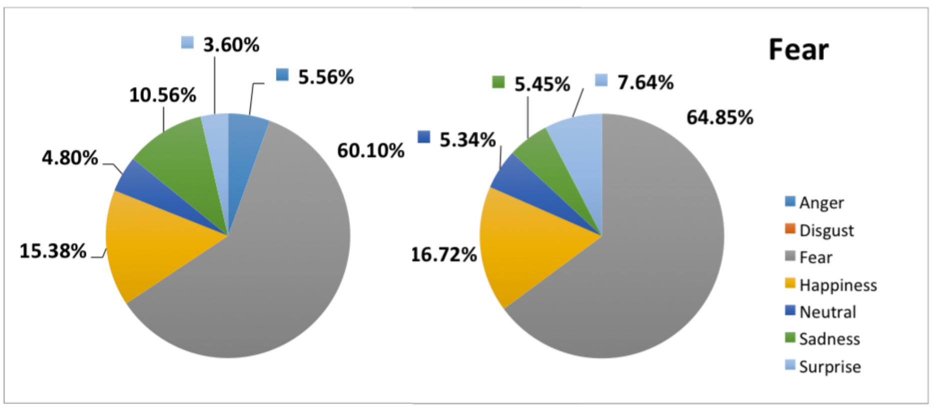


Figure 4.28: The comparison of fear estimation of PCA & EMC between with eyeglasses and with reducing the effect of eyeglasses

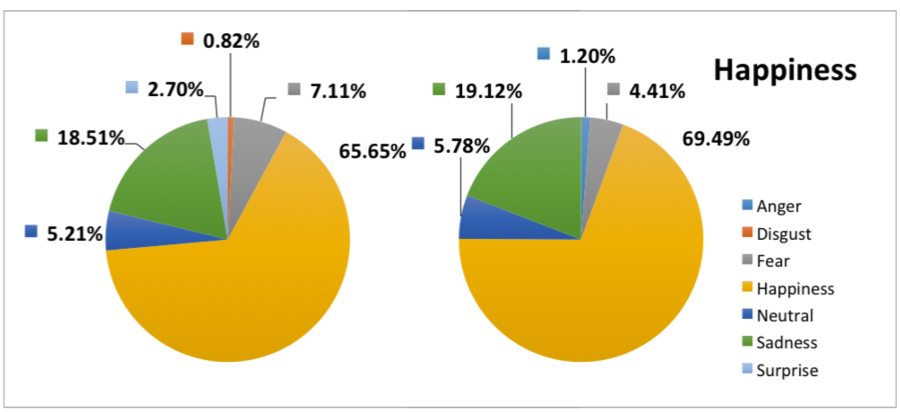


Figure 4.29: The comparison of happiness estimation of PCA & EMC between with eyeglasses and with reducing the effect of eyeglasses

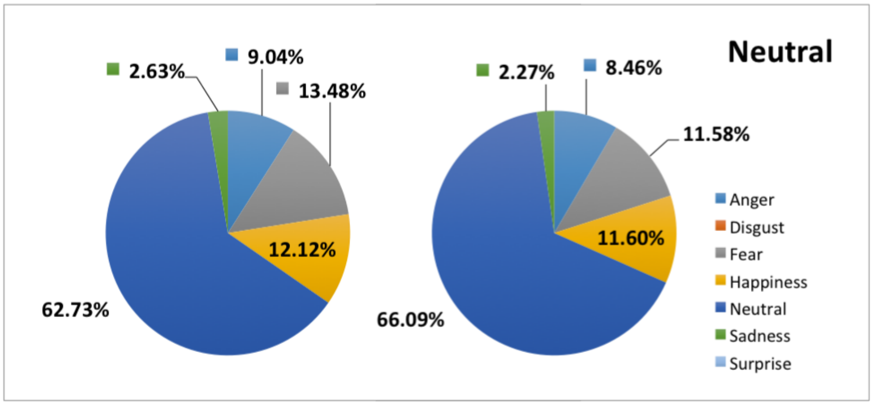


Figure 4.30: The comparison of neutral estimation of PCA & EMC between with eyeglasses and with reducing the effect of eyeglasses

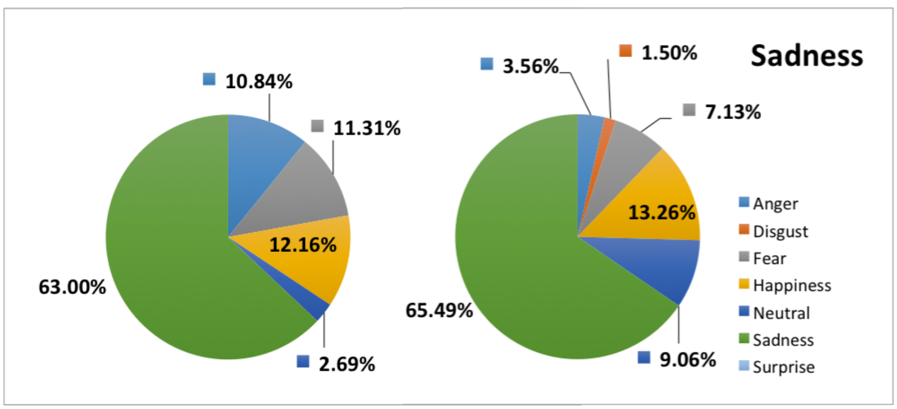


Figure 4.31: The comparison of sadness estimation of PCA & EMC between with eyeglasses and with reducing the effect of eyeglasses

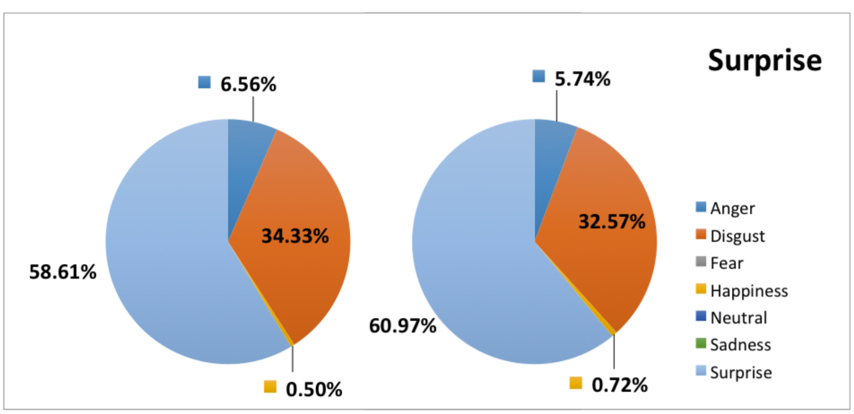


Figure 4.32: The comparison of surprise estimation of PCA & EMC between with eyeglasses and with reducing the effect of eyeglasses

Chapter 5

Estimation of Human Emotion Using t-ROI for Thermal IR Image Sequence

Human emotion estimation is a rapidly growing area of research due to its increasing implications on the development of real human-computer interaction (HCI) systems. Most of the studies in this field use visible image-based representations to estimate human emotions. However, under uncontrolled operating conditions, estimation accuracy degrades significantly. Furthermore, some people have poker-faces, whereas others have their expressions differing from their true emotions. Therefore, using thermal IR images and thermal IR information gives several advantages over conventional visible images because of its invariance to illumination changes. Nevertheless, IR imagery has several drawbacks, such as being opaque to eyeglasses. Besides, there exist several dominant facial parts which impact to the emotion change. To address this serious limitation of IR imagery, we propose using Thermal Regions of Interest (t-ROI). We also use image sequences to estimate human emotion because a static image can not give a true emotion. For this experiment, the Kotani Thermal Facial Emotion (KTFE) database is used.

5.1 Introduction

In the last decade, automated estimation of human emotion has attracted the interest of many researchers, because such systems will have numerous applications in security, medicine, and especially human-computer interaction. Many previous works proposed have been inclined towards developing facial emotion estimation [96]. Nevertheless, there is a lack of accurate and robust facial emotion estimation methods to be deployed in uncontrolled environments. When the lighting is dim or when it does not uniformly illuminate the face, the accuracy decreases considerably. Moreover, human emotion estimation based on only the visible spectrum has proved to be difficult in cases where there are emotion changes that expressions do not show. Using thermal IR image is a new and innovative way which is not sensitive to light conditions, to fill the gap in the human emotion estimation field. Besides, human emotions could be manifested by changing temperature of face skin which is obtained by an IR camera. Consequently, thermal IR image gives us more information to help us robustly estimate human emotions.

Although there are many significant advantages when we use IR imagery, it has several drawbacks. Firstly, thermal IR data are subjected to change together with body temperature caused by variable ambient temperatures. Secondly, presence of eyeglasses may result in loss of useful information around the eyes. Glass is opaque to IR, and object made of glass act as temperature screen, completely occluding the parts located behind them. Hence, the sensitivity of IR imagery is decreased by facial occlusions. Thirdly, there are some facial regions not receptive to the emotion changes. To eliminate the effects of the three challenging problems above, we propose using Thermal Region of Interest (t-ROI) for thermal IR data. We will also reduce the impact of ambient temperature by normalizing it between different frames. Furthers, single images give only spatial information while image sequences contain both the temporal and spatial structure of the phenomena observed. And static image can not give the true emotion. Therefore, we propose t-ROIs for image sequence. To estimate seven emotions, we use three our proposed methods Principal Component Analysis (PCA), Eigen-space Method based on class-features (EMC), and PCA-EMC and two proposed methods, t-PCA (thermal PCA) and n-EMC (norm EMC) over obtained t-ROIs to extract feature vectors and then find the similarity

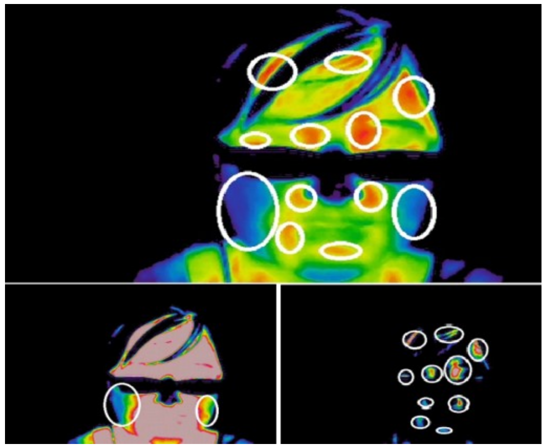


Figure 5.1: An example of t-ROIs.

between the testing data and training data.

5.2 Methods

Before selecting features, we perform some preprocessings such as face normalization, noise deduction.

First, with sequences of thermal IR images, we find the regions of interest based on t-ROIs.
In our definition, interest regions are regions in which temperature increases or decreases significantly when human emotions change. We use the two regions which are the hottest and coldest regions of the face, except the eyeglasses, usually the forehead, eyeholes, and cheek-bone regions, as our interest regions. Before finding the t-ROIs, to avoid any ambient temperature change from frame to frame, we update the temperature of each point of each frame based on the difference between mean of ambient temperature and mean of the first m frame ambient temperature.
Let h be a map from face ($Fa \subset R^2$) to temperature ($T \subset R$) space

$$h : Fa \to T$$

$$(i, j) \mapsto h(i, j)$$

We obtain the t-ROIs by using the following equations:

$$\Delta T_{Fa} = T^{Fa}_{Max} - T^{Fa}_{Min}; \delta T_{Fa} = \Delta T_{Fa}/5$$

$$L^{Fa}_{k,idx} = \{(i, j) \in Fa | T^{Fa}_{Min} + \delta T_{Fa} * (idx - 1) \leq h(i, j) < T^{Fa}_{Max} - \delta T_{Fa} * (5 - idx)\} \quad (5.1)$$

$$L_i dx = L^{Fa}_{k,idx}, k \in \overline{1, K}$$

where $T^{Fa}_{Max}, T^{Fa}_{Min}$ are maximum and minimum of temperature of each human face at frame k, respectively; K is the number of frame; $idx \in \{2, 5\}$.

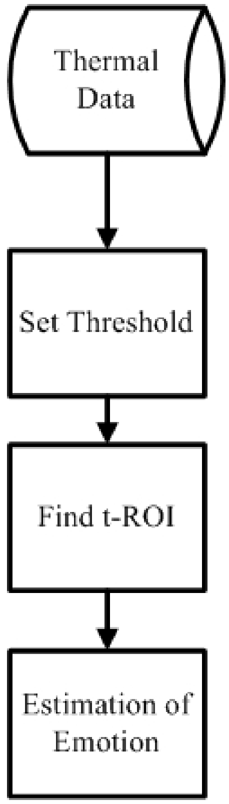


Figure 5.2: t-ROIs for human emotion estimation.

To estimate the true emotion, using single frame can not give the exact emotion, hence we use the sequence of thermal IR frames. Therefore, after reducing the effect of eyeglasses, to apply for sequence of thermal IR frames, we calculate the accumulate of the discriminant of frame m and frame m + idx.

$$L_i dx : F^* \to T$$
$$(i,j) \mapsto f^*(i,j)$$

We calculate $\sum_m \parallel f^*_m(i,j) - f^*_{m+idx}(i,j) \parallel, \forall (i,j) \in L_i dx$

After obtaining the t-ROIs for each frame, we find the dominant levels of frames. A frame a is more dominant than frame b if only if temperature change of all t-ROIs between frame a and frame a-k is bigger than temperature change of all t-ROIs between frame b and frame b-k. Based on the obtained dominant level of each frame, we can automatically put the weight for each frame.

Let call F is the number of frames. $\forall f_m \in F$

$$w_{f_m} > w_{f_n} \Longleftrightarrow \sum_{\forall i,j \in L_i dx)} \parallel f^*_m(i,j) - f^*_{m+idx}(i,j) \parallel \geq \sum_{\forall i,j \in L_i dx)} \parallel f^*_n(i,j) - f^*_{n+idx}(i,j) \parallel,$$

To illustrate the feasibility of using eigen-space to fulfill facial emotion estimation task, thermal PCA (t-PCA) is modified from the PCA reconstruction method and evaluated over thermal IR data [113]. With PCA, the aim is to build a face space, including the

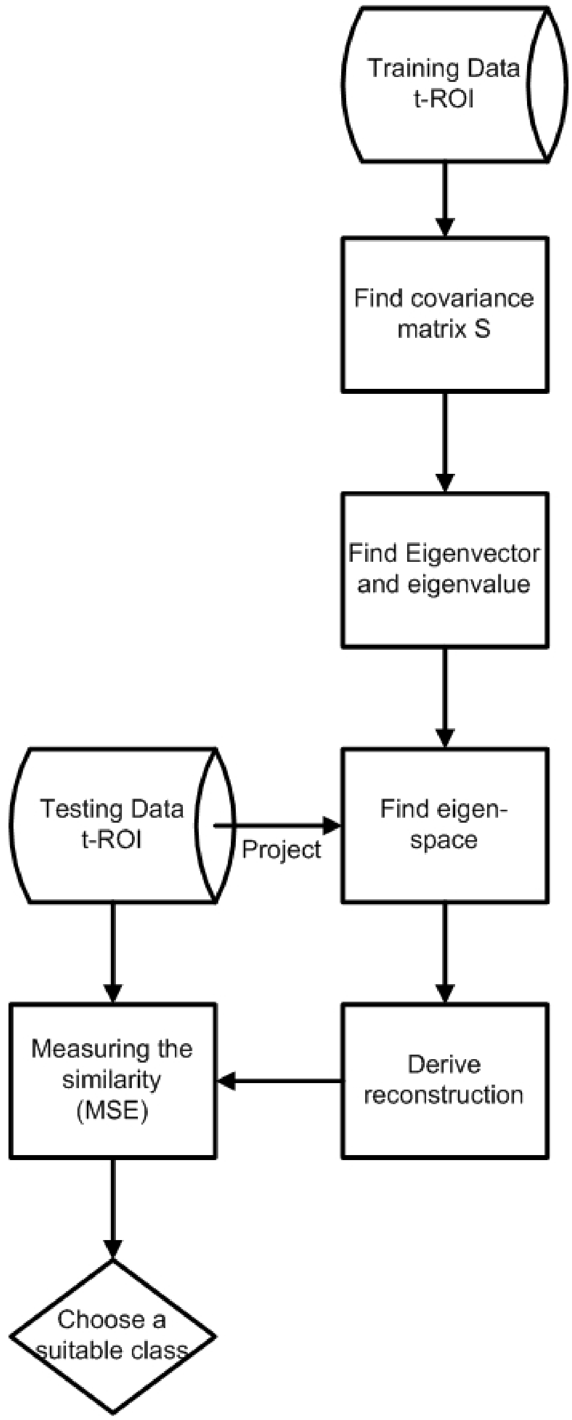


Figure 5.3: PCA for human emotion estimation.

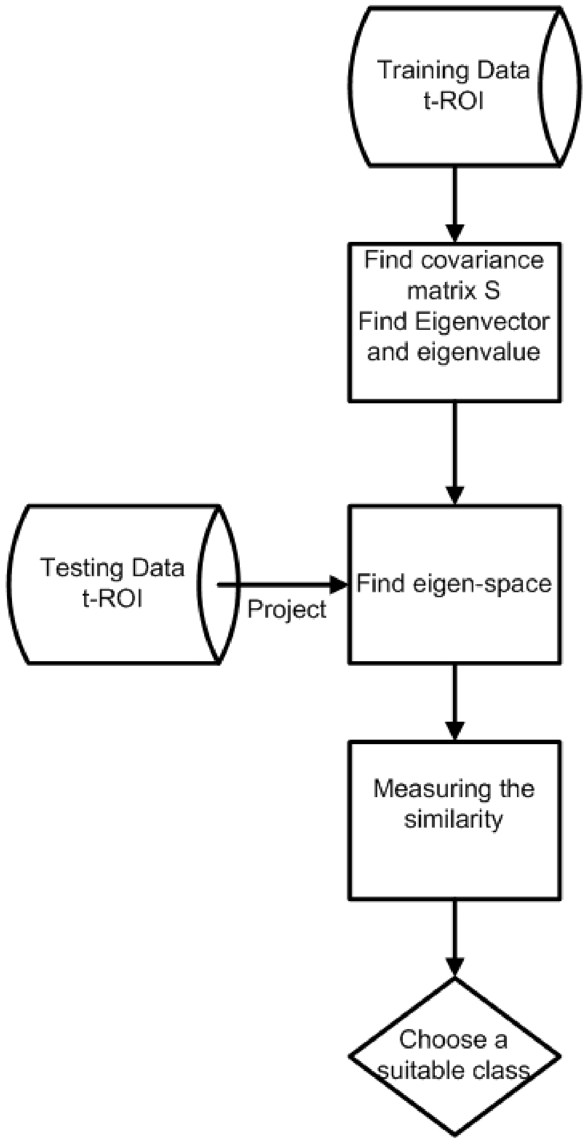


Figure 5.4: EMC for human emotion estimation.

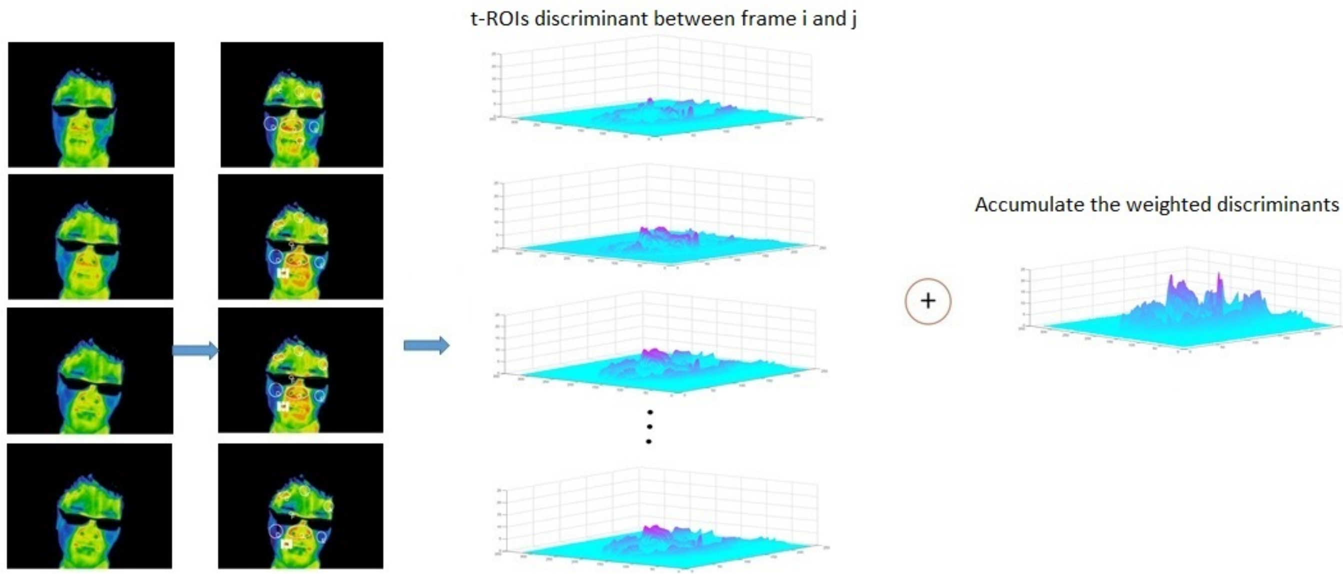


Figure 5.5: Accumulate the weighted discriminants of t-ROIs.

basis vectors called principal components, which better describe the face images. PCA has several advantages over other face recognition schemes in its speed and simplicity [113]. We modified PCA to estimate facial emotion from thermal IR data.

Let F be a set of classes to be analyzed. Here, F is a set of all emotion classes. Assume that M_m^f thermal IR frames of training data are given as the facial temperature pattern for each class $f \in F$ where $F = \{anger, disgust, fear, happiness, neutral, sadness, surprise\}$. Let Γ_i^f be the $i - th$ facial temperature pattern where $i = \overline{1, M_f}$; the dimension of Γ_i^f, $n \times m$, is equal to the number of pixels in a thermal IR frame, and each element of Γ_i^f indicates the temperature of each pixel.

Compute the mean of training data $\Psi^f = \frac{1}{M^f} \sum_{i=1}^{M_f} \Gamma_i^f$ and let the normalized vector be $\Phi_i^f = \Gamma_i^f - \Psi^f$. We seek a set of M orthonormal vectors, u_k^f, which best describes the distribution of the training data. The kth vector, u_k^f is chosen by

$$\lambda_k^f = \frac{1}{M_f} \sum_{i=1}^{M_f} ((u_k^f)^\tau \Phi_i^f)^2. \tag{5.2}$$

is a maximum, subject to $(u_l^f)^\tau u_k^f = \begin{cases} 1, & \text{if } l = k. \\ 0, & \text{if } otherwise. \end{cases}$

The eigenvectors and eigenvalues are the vector u_k^f and scalar λ_k^f of covariance matrix $C^f = \frac{1}{M_f} \sum_{i=1}^{M_f} (\Phi_i^f (\Phi_i^f)^\tau = A^f (A^f)^\tau$ where $A^f = [\Phi_1^f, \Phi_2^f, ..., \Phi_{M_f}^f]$

The size of covariance matrix C^f, $nm \times nm$, is too large to determine $n \times m$ eigenvectors and eigenvalues. Turk et al. [118] suggested a computationally feasible method to find these eigenvectors. Let a matrix $H^f = (A^f)^\tau A^f$, then size of H^f is size $M \times M \ll nm \times nm$; let ν_i^f denote the eigenvectors of H^f.

We have $(A^f)^\tau A^f \nu_i^f = \mu_i^f \nu_i^f \Leftrightarrow A^f (A^f)^\tau A^f \nu_i^f = A^f \mu_i^f \nu_i^f \Leftrightarrow A^f (A^f)^\tau A^f \nu_i^f = A^f \mu_i^f \nu_i^f$

$$\Leftrightarrow C^f A^f \nu_i^f = \mu_i^f A^f \nu_i^f \tag{5.3}$$

From Equation(4), $A^f \nu_i^f$ is the eigenvector of $C^f = A^f (A^f)^\tau$ [118]. Therefore we can obtain ρ^f eigenvector of C^f by calculating the $\rho^f(\rho^f \ll nm)$ eigenvectors (ν_i^f) of matrix H^f and multiplying A^f to ν_i^f. After obtaining eigenfaces from training data of each emotion, we map facial thermal IR training data to feature spaces by $\omega_i^f(train) = (u_i^f)^\tau(\Gamma^f - \Psi^f), i = \overline{1, \rho^f}$.

We use the idea that if the input frame is much similar to some emotion training set, the reconstructed data will has less distortion than the data reconstructed from other eigenvectors of training emotions [113]. For each testing facial thermal IR frame Γ_{test}, firstly we project it onto the eigenfaces of each class.

$\omega_{test}^f = (U^f)^\tau(\Gamma_{test} - \Psi^f)$, where $U^f = (u_i^f), i = \overline{1, \rho^f}$.

Secondly, for each emotion, we find the feature which is most similar to the testing projected vector by calculate the angle between vector of training feature space and testing projected vector.

$$\beta^f = \operatorname*{argmax}_i \frac{\omega_{test}^f \omega_i^f(train)}{\|\omega_{test}^f\| \|\omega_i^f(train)\|}; i = \overline{1, \rho^f}. \tag{5.4}$$

Thirdly, we find reconstruction of the testing data by the obtained feature in each class.

$\Gamma_{reconst}^f = U^f \omega_{\beta^f}^f + \Psi^f$

Finally, we choose an emotion of which reconstruction of the testing data is the most similarity to testing data.

$$\gamma = \operatorname*{argmax}_f \frac{\Gamma_{test}\Gamma_{reconst}^f}{\|\Gamma_{test}\| \|\Gamma_{reconst}^f\|}; f = \overline{1, 7} \tag{5.5}$$

The second valuation to estimate human emotion uses n-EMC over thermal IR data. n-EMC is modified from EMC [114]. The difference between EMC and n-EMC is formulation to calculate the difference between the within-class and between-class variance.

In mathematics, with n-EMC, instead of finding the eigenvectors, u_k^f and eigenvalues λ_k^f of covariance matrix $C^f = \frac{1}{M_f}\sum_{i=1}^{M_f}(\Phi_i^f(\Phi_i^f)^\tau = A^f(A^f)^\tau$ where $A^f = [\Phi_1^f, \Phi_2^f, ..., \Phi_{M_f}^f]$, we find eigenvectors, u_k^f and eigenvalues λ_k^f of matrix $S = \|S_B - S_W\|_2$ where

$$M = \sum_{f \in F} M_f \tag{5.6}$$

$$\Psi^f = \frac{1}{M^f}\sum_{i=1}^{M_f}\Gamma_i^f; \Psi = \frac{1}{M}\sum_{f \in F}\sum_{i=1}^{M_f}\Gamma_i^f \tag{5.7}$$

$$S_B = \frac{1}{M}\sum_{f \in F} M_f \|\Psi_f - \Psi\|_2 \|\Psi_f - \Psi\|_2^\tau. \tag{5.8}$$

$$S_W = \frac{1}{M}\sum_{f \in F}\sum_{i=1}^{M_f}\|\Gamma_i^f - \Psi_f\|_2 \|\Gamma_i^f - \Psi_f\|_2^\tau. \tag{5.9}$$

For each testing facial thermal IR frame Γ_{test}, firstly we project it onto the eigenfaces of each class.

$\omega_{test}^f = (U^f)^\tau(\Gamma_{test} - \Psi^f)$, where $U^f = (u_i^f), i = \overline{1, \rho^f}$.

Secondly, for each emotion, we find the feature which is most similar to the testing projected vector by calculate the angle between vector of training feature space and testing projected vector.

$$\beta^f = \underset{i}{\operatorname{argmax}} \frac{\omega^f_{test}\omega^f_i(train)}{\|\omega^f_{test}\|\|\omega^f_i(train)\|}; i = \overline{1, \rho^f}. \tag{5.10}$$

Finally, we choose an emotion which has maximum of β^f

$$\gamma = \underset{f}{\operatorname{argmax}} \beta^f; f = \overline{1, 7} \tag{5.11}$$

5.3 Experiments

In our experiments, we use only sequence of thermal IR data to estimate human emotions. From 186 GB thermal data, we extracted 28 GB of sequence thermal IR data for seven emotions. We separate the training and testing data as 60% and 40% of total sequence of thermal IR images.

Table.5.1 shows the results of facial emotion estimation of PCA for sequence of thermal IR images. The average of accuracy using t-ROIs for sequence of thermal IR images increases 6.79% in comparison with without using t-ROIs. Especially, for the anger emotion, the average of accuracy increase 9.96 % with using t-ROIs.

Table.5.2 shows the results of facial emotion estimation of t-PCA for sequence of thermal IR images. The average of accuracy using t-PCA for t-ROIs for sequence of thermal IR images increases 1.65% in comparison with PCA for t-ROIs. Especially, for the happiness emotion, the average of accuracy increase of 3.02% with using t-ROIs.

Table.5.3 shows the results of facial emotion estimation of EMC for sequence of thermal IR images. The average of accuracy using t-ROIs for sequence of thermal IR images increases 3.91% in comparison with without using t-ROIs. Especially, for the surprise emotion, the average of accuracy increase 6.01 % with using t-ROIs.

Table.5.4 shows the results of facial emotion estimation of n-EMC for sequence of thermal IR images. The average of accuracy using n-EMC for t-ROIs for sequence of thermal IR images increases 0.75% in comparison with EMC for t-ROIs. Especially, for the neutral emotion, the average of accuracy increase 1.5% with using t-ROIs.

Table.5.5 shows the results of facial emotion estimation of PCA & EMC for sequence of thermal IR images. The average of accuracy using t-ROIs for sequence of thermal IR images increases 5.16% in comparison with without using t-ROIs. Especially, for the fear emotion, the average of accuracy increase 11.94 % with using t-ROIs.

5.4 Conclusions

In this chapter, we propose a novel way, t-ROI for sequence of thermal IR images, to eliminate the drawbacks of IR imagery in human emotion estimation. Specially, two classification methods, t-PCA and n-EMC, are proposed. Our method has several advantaged points. First, to the best of our knowledge, this is one of the first methods using sequences of thermal IR images. Emotion is complex action of human. To understand it clearly, using a single image cannot figure out the exact emotion. Besides, using thermal

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	85.59	0.53	4.32	3.38	-	0.53	5.64
Disgust	-	84.89	3.62	-	-	-	11.49
Fear	2.64	0.01	84.22	6.22	2.34	4.64	0.01
Happiness	-	0.14	2.50	83.69	1.25	12.42	-
Neutral	0.56	-	2.97	4.69	83.31	8.48	-
Sadness	1.68	-	3.26	3.33	5.65	86.08	-
Surprise	11.10	6.14	-	1.00	-	-	81.76
Average							**84.22**

Table 5.1: The confusion matrix of PCA with t-ROIs

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	86.47	1.64	2.78	3.44	-	1.64	4.05
Disgust	-	86.70	2.77	-	-	-	10.53
Fear	3.83	0.01	85.35	3.44	2.91	4.54	0.01
Happiness	-	0.28	0.46	86.71	0.23	12.33	-
Neutral	4.63	-	2.52	3.69	85.66	3.50	-
Sadness	-	0.84	3.33	6.72	0.27	88.85	-
Surprise	2.51	15.32	-	0.83	-	-	81.35
Average							**85.87**

Table 5.2: The confusion matrix of tPCA with t-ROIs

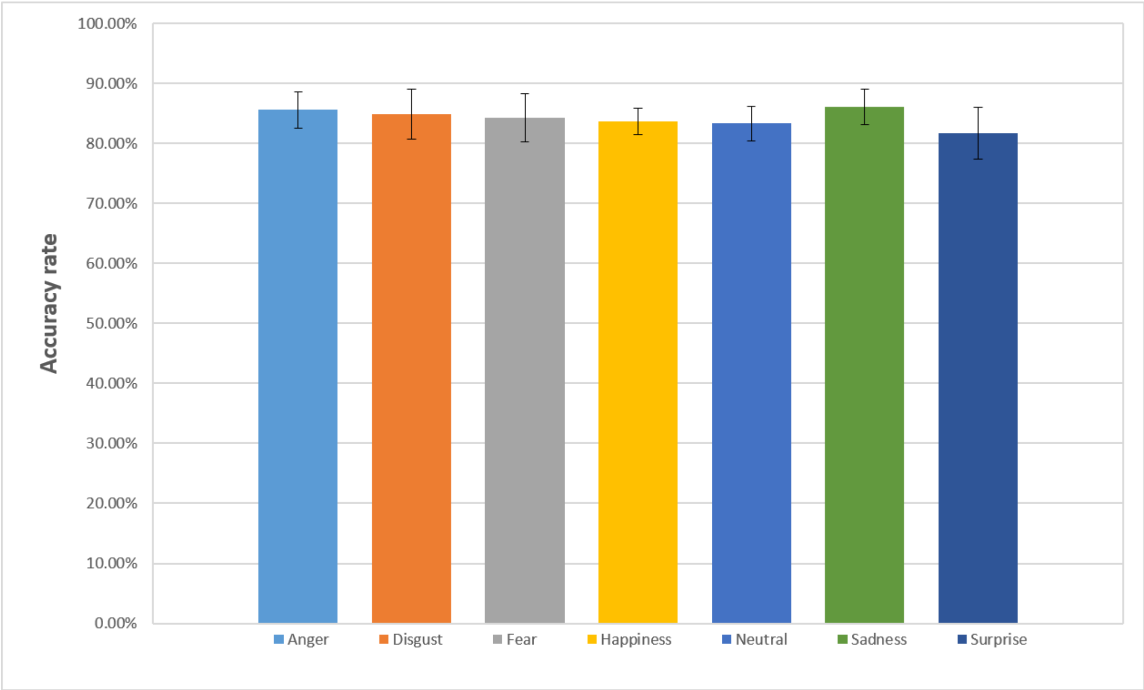


Figure 5.6: The emotion estimation results of PCA with t-ROIs

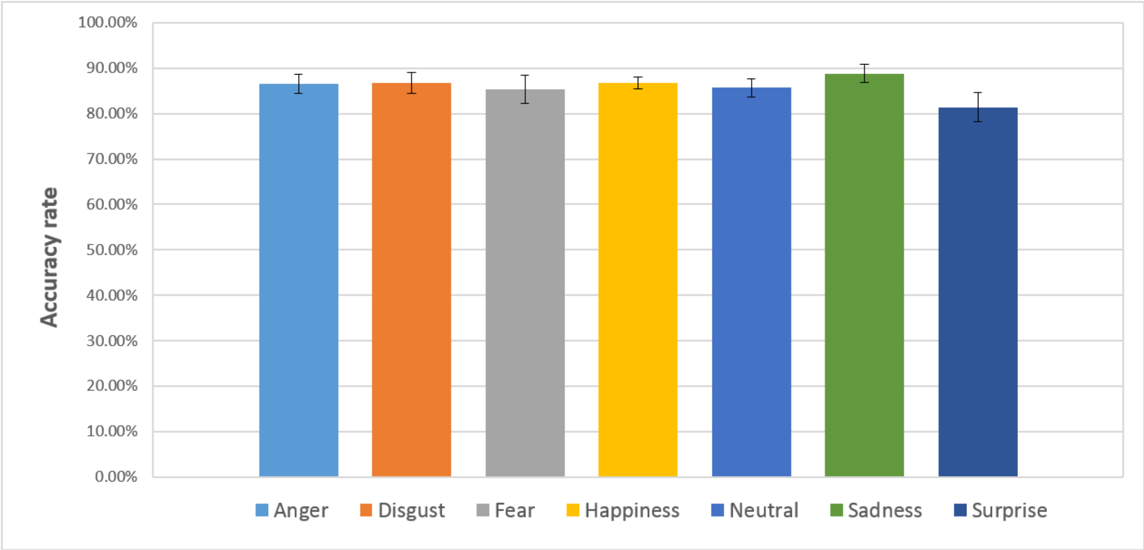


Figure 5.7: The emotion estimation results of t-PCA with t-ROIs

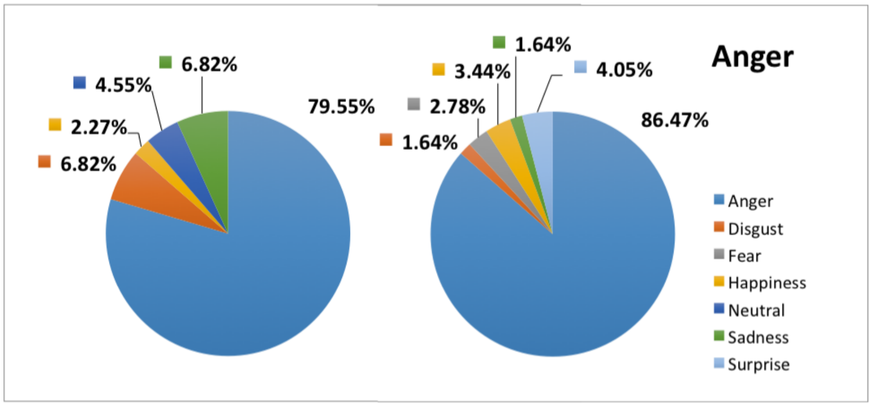


Figure 5.8: The comparison of anger estimation of t-PCA between without using t-ROIs and using t-ROIs

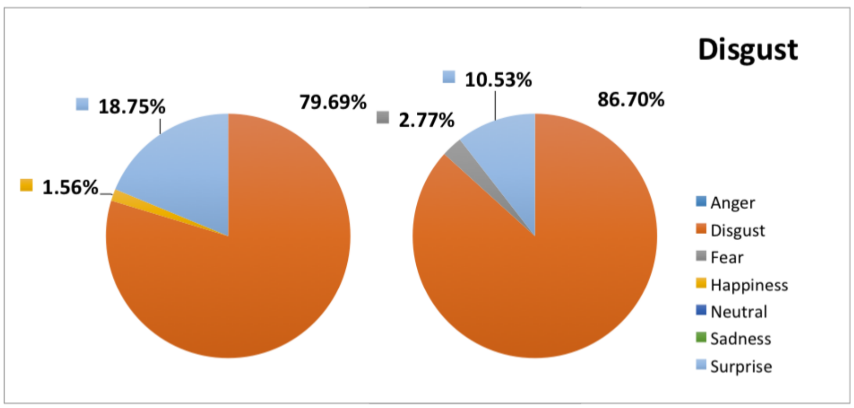


Figure 5.9: The comparison of disgust estimation of t-PCA between without using t-ROIs and using t-ROIs

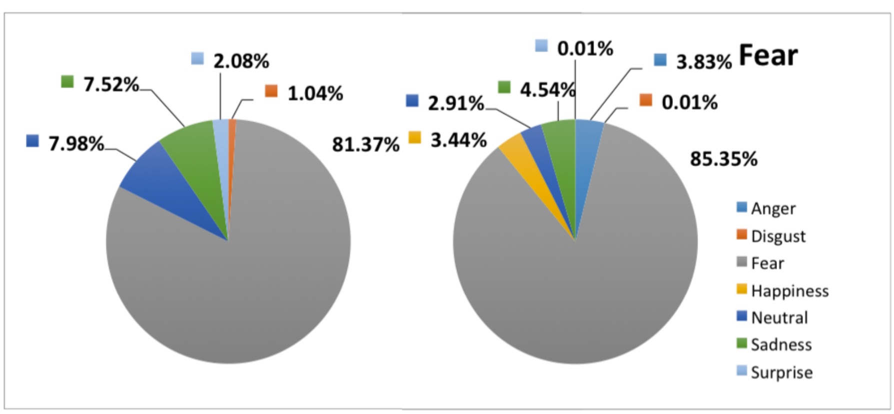


Figure 5.10: The comparison of fear estimation of t-PCA between without using t-ROIs and using t-ROIs

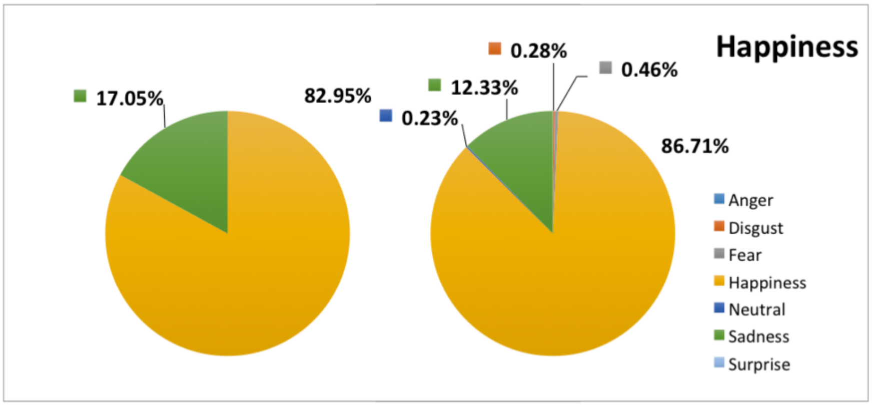


Figure 5.11: The comparison of happiness estimation of t-PCA between without using t-ROIs and using t-ROIs

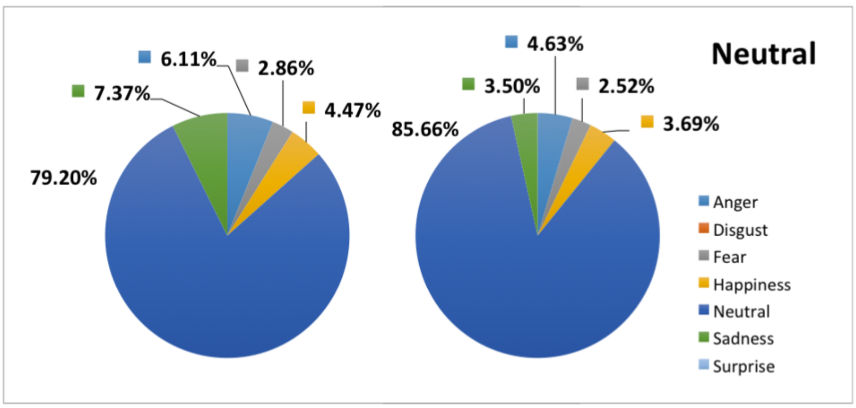


Figure 5.12: The comparison of neutral estimation of t-PCA between without using t-ROIs and using t-ROIs

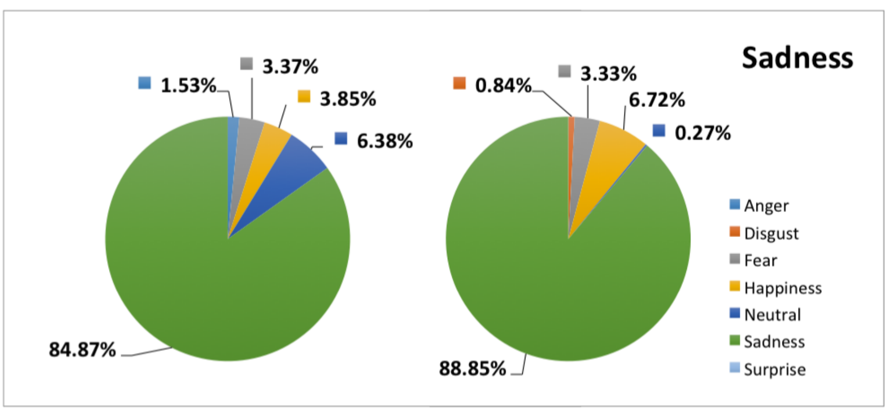


Figure 5.13: The comparison of sadness estimation of t-PCA between without using t-ROIs and using t-ROIs

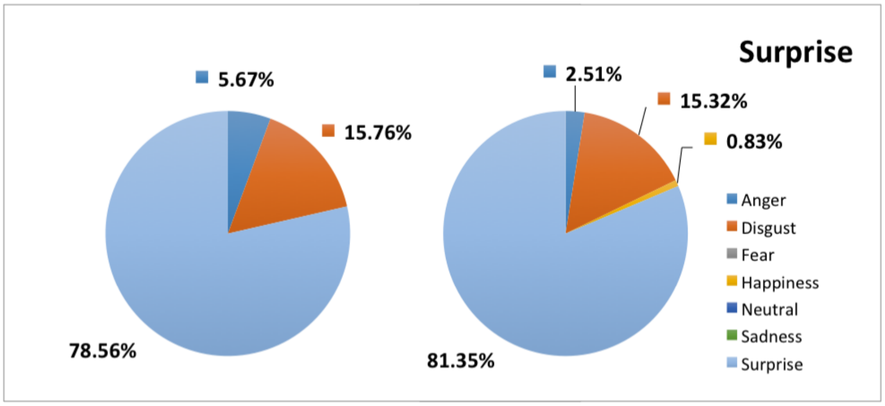


Figure 5.14: The comparison of surprise estimation of t-PCA between without using t-ROIs and using t-ROIs

IR information with single frame cannot give the right emotion. Therefore, it is necessary to use sequences of thermal IR images. Second, with t-ROIs, we fill the gaps of thermal IR image, eyeglass problem. Third, using the weight discriminant features help our method decrease the running cost and increase accuracy rate. Because there are some frames more important than others, the weights set for frames are necessary. Four, two proposed methods, t-PCA and n-EMC, work better than PCA and EMC, respectively in our database for human emotion estimation. The experiment results show the average accuracy of estimating emotion using t-ROIs is better than the accuracy of estimation emotion without using t-ROIs.

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	86.36	6.82	-	6.82	-	-	-
Disgust	-	82.22	5.56	-	-	-	12.22
Fear	3.06	-	81.99	6.00	1.17	7.79	-
Happiness	-	-	-	87.98	-	12.02	-
Neutral	14.57	-	-	-	84.57	0.86	-
Sadness	-	-	16.35	-	1.78	81.88	-
Surprise	-	8.83	-	-	-	8.83	82.33
Average							**83.90**

Table 5.3: The confusion matrix of EMC with t-ROIs

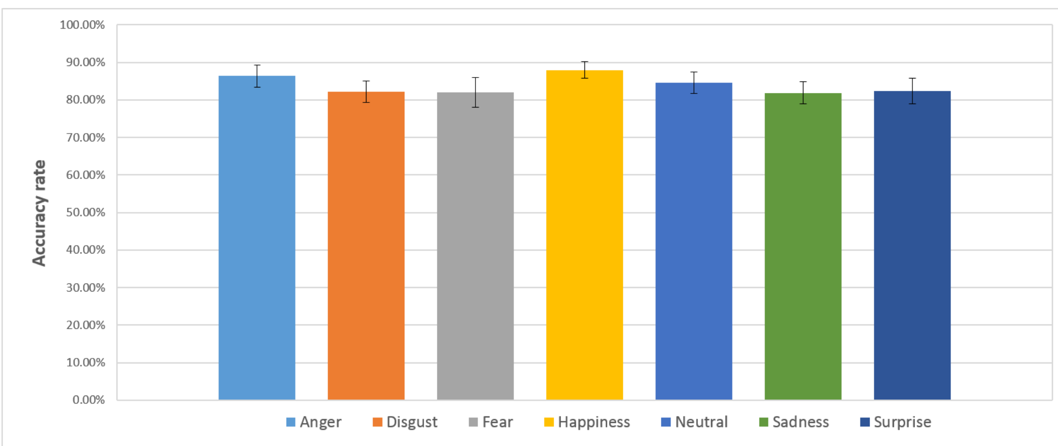


Figure 5.15: The emotion estimation results of EMC with t-ROIs

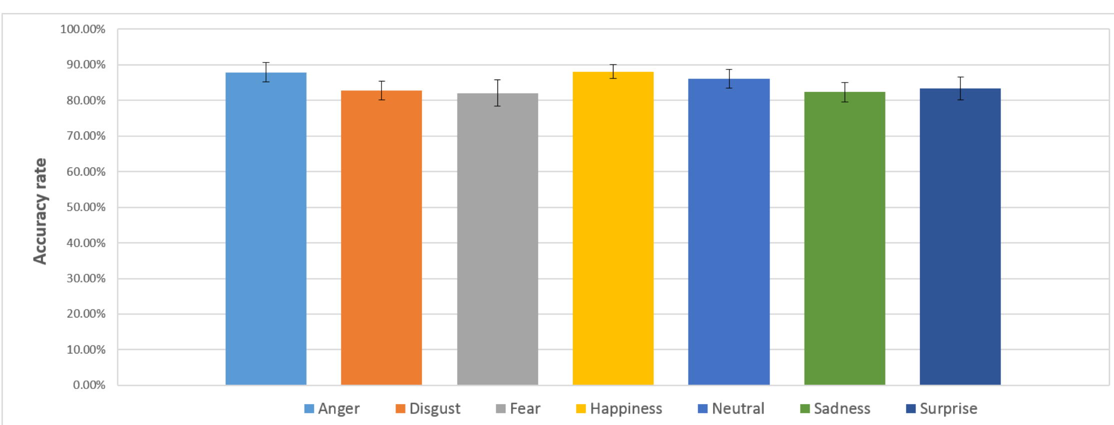


Figure 5.16: The emotion estimation results of n-EMC with t-ROIs

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	87.84	5.46	-	5.45	-	1.25	-
Disgust	-	82.79	1.67	-	-	-	15.54
Fear	-	-	82.07	11.54	1.39	5.00	-
Happiness	-	-	-	88.06	-	11.94	-
Neutral	12.86	-	-	-	86.07	1.07	-
Sadness	-	-	16.12	-	1.58	82.31	-
Surprise	-	8.28	-	-	-	8.28	83.44
Average							**84.65**

Table 5.4: The confusion matrix of n-EMC with t-ROIs

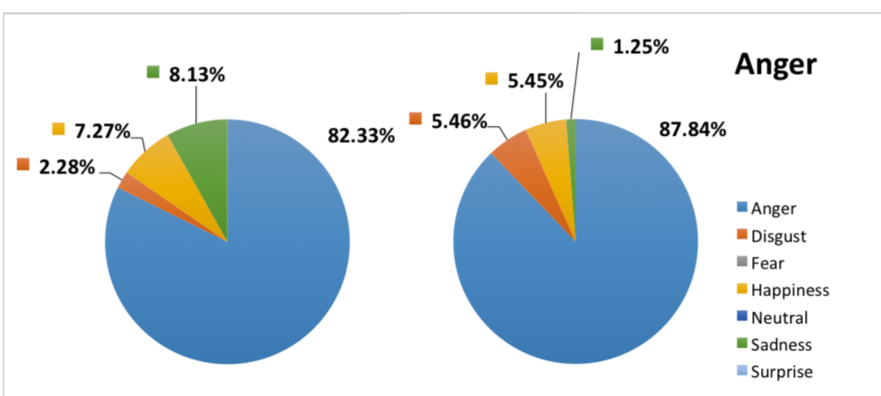


Figure 5.17: The comparison of anger estimation of n-EMC between without using t-ROIs and using t-ROIs

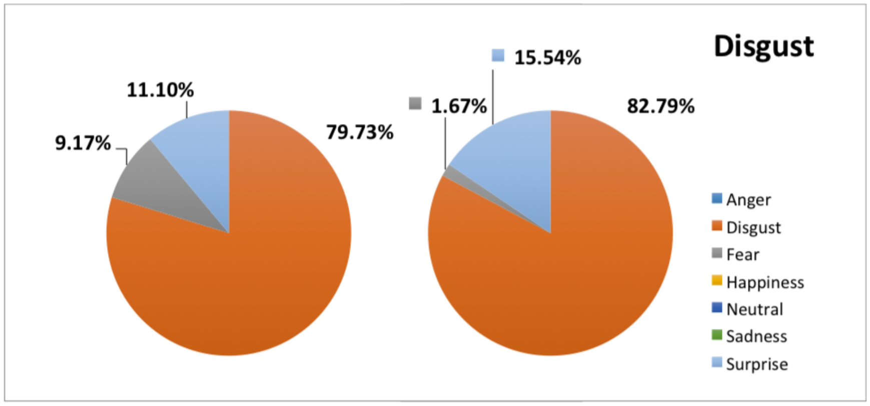


Figure 5.18: The comparison of disgust estimation of n-EMC between without using t-ROIs and using t-ROIs

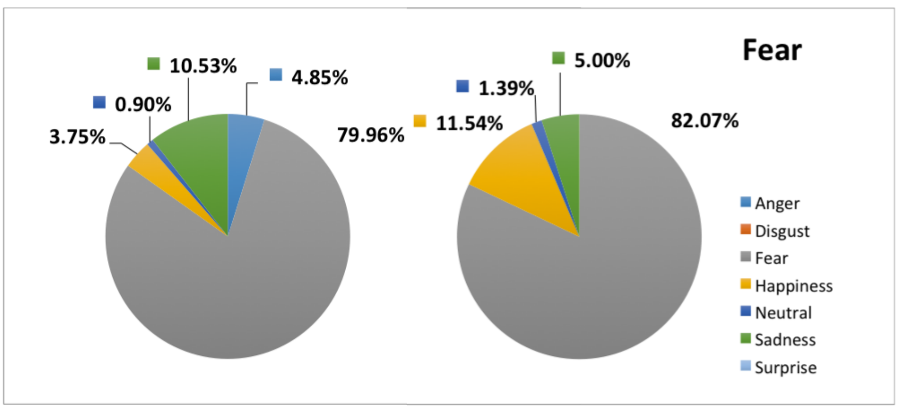


Figure 5.19: The comparison of fear estimation of n-EMC between without using t-ROIs and using t-ROIs

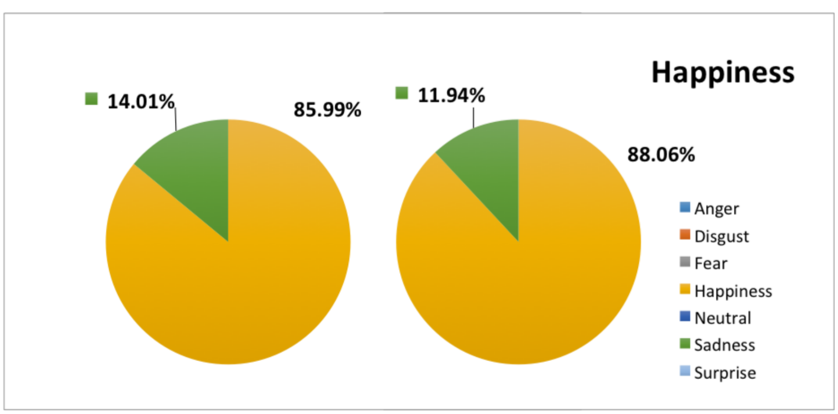


Figure 5.20: The comparison of happiness estimation of n-EMC between without using t-ROIs and using t-ROIs

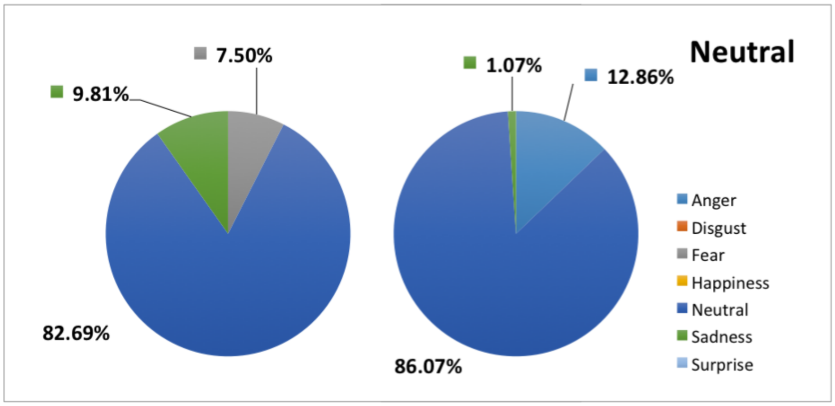


Figure 5.21: The comparison of neutral estimation of n-EMC between without using t-ROIs and using t-ROIs

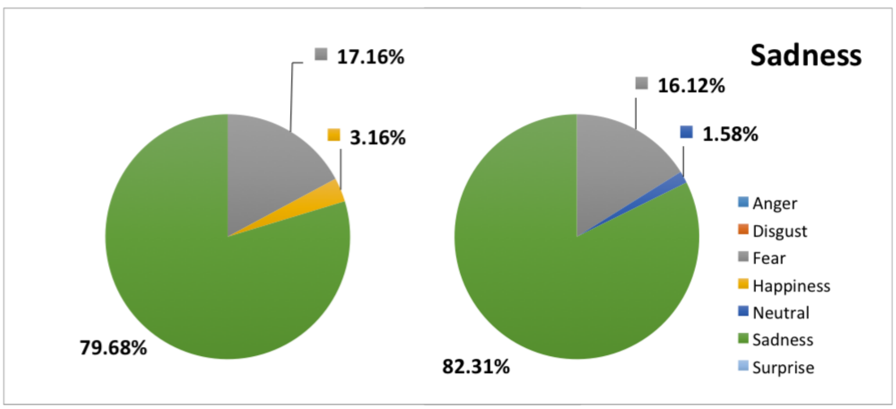


Figure 5.22: The comparison of sadness estimation of n-EMC between without using t-ROIs and using t-ROIs

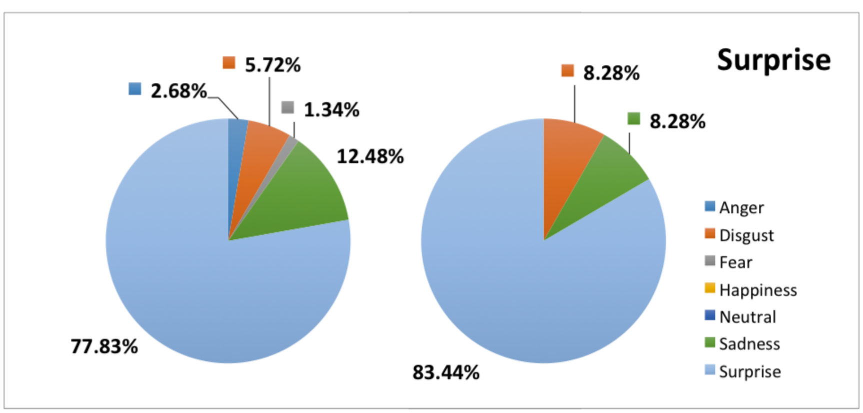


Figure 5.23: The comparison of surprise estimation of n-EMC between without using t-ROIs and using t-ROIs

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	64.13	8.53	9.58	4.55	-	10.46	2.75
Disgust	-	70.55	6.25	4.36	-	-	18.84
Fear	-	-	66.74	16.29	5.32	3.58	8.07
Happiness	1.20	-	4.40	69.79	5.65	18.96	-
Neutral	9.10	-	11.21	9.98	67.96	1.75	-
Sadness	4.34	1.20	6.06	12.95	8.87	66.58	-
Surprise	5.22	32.20	-	0.66	-	-	61.92
Average							**66.81**

Table 5.5: The confusion matrix of PCA & EMC with t-ROIs

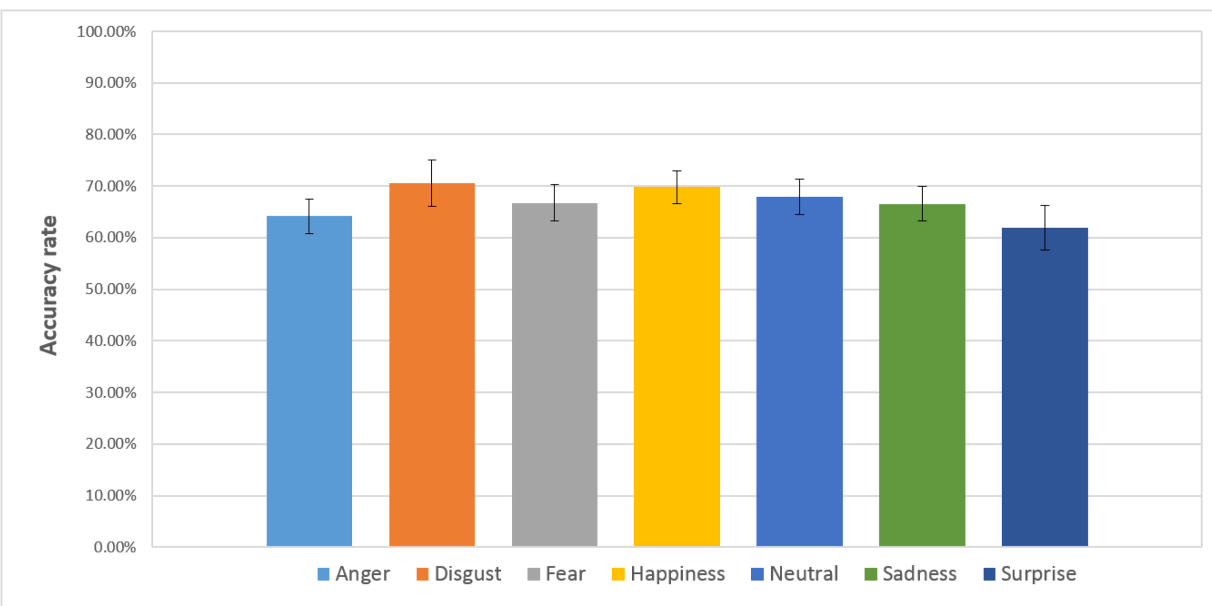


Figure 5.24: The emotion estimation results of PCA & EMC with t-ROIs

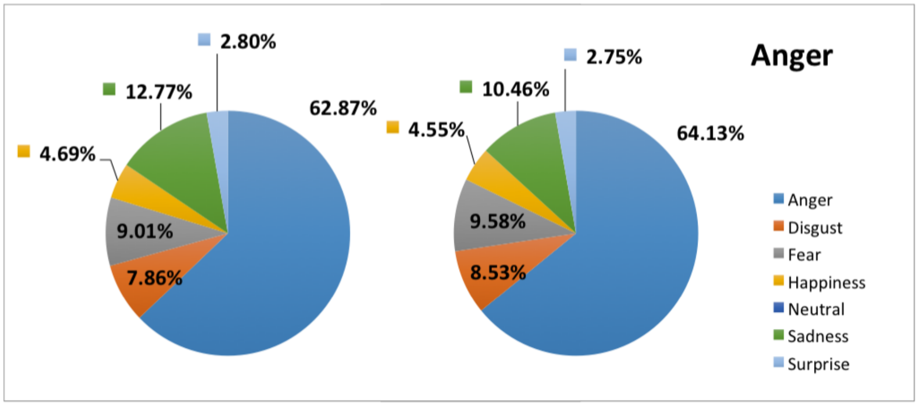


Figure 5.25: The comparison of anger estimation of PCA & EMC between without using t-ROIs and using t-ROIs

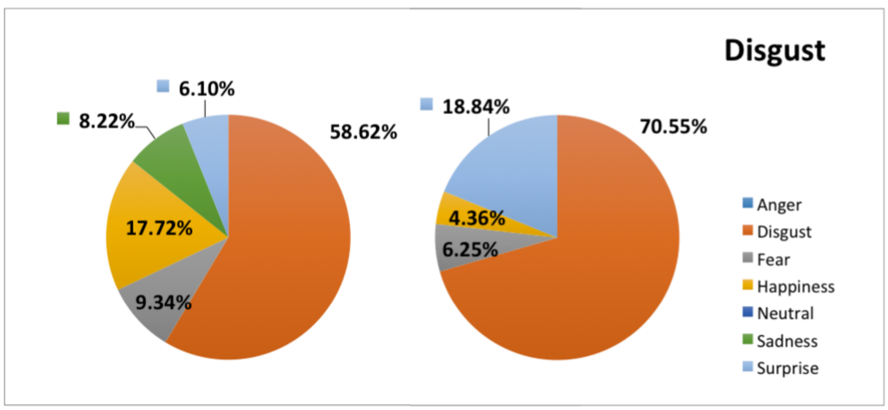


Figure 5.26: The comparison of disgust estimation of PCA & EMC between without using t-ROIs and using t-ROIs

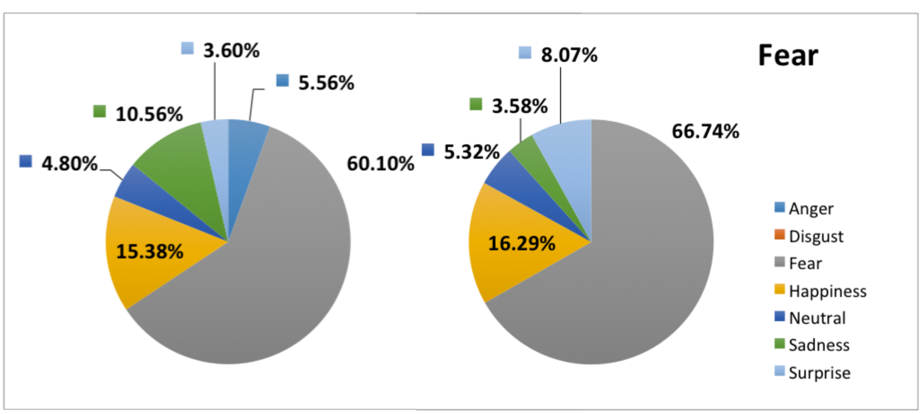


Figure 5.27: The comparison of fear estimation of PCA & EMC between without using t-ROIs and using t-ROIs

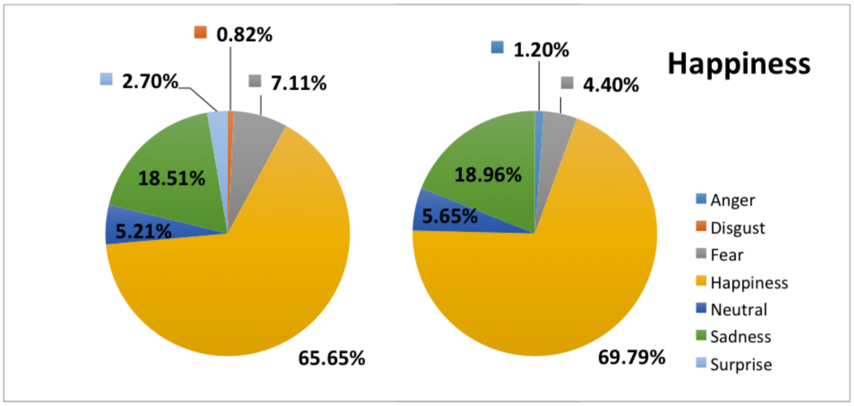


Figure 5.28: The comparison of happiness estimation of PCA & EMC between without using t-ROIs and using t-ROIs

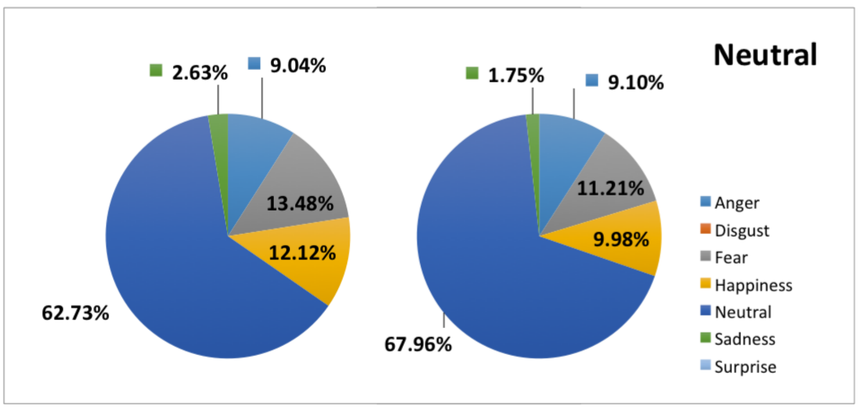


Figure 5.29: The comparison of neutral estimation of PCA & EMC between without using t-ROIs and using t-ROIs

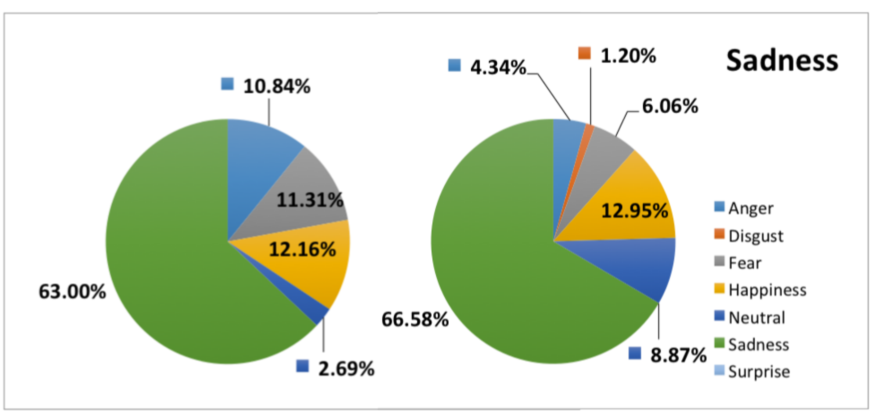


Figure 5.30: The comparison of sadness estimation of PCA & EMC between without using t-ROIs and using t-ROIs

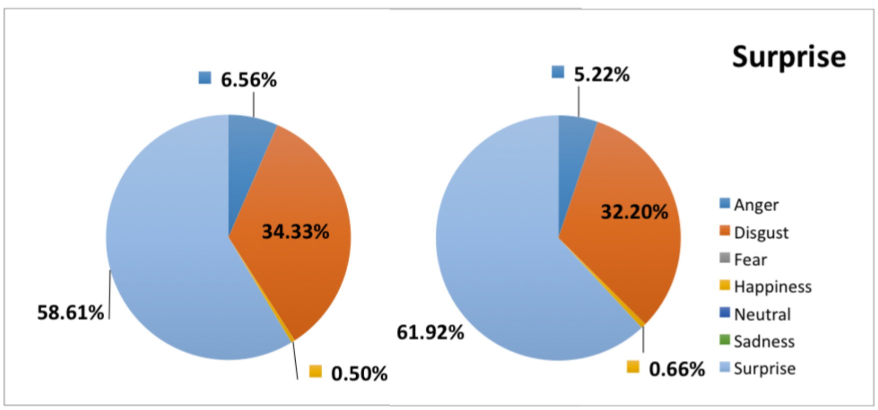


Figure 5.31: The comparison of surprise estimation of PCA & EMC between without using t-ROIs and using t-ROIs

Chapter 6

Estimation of Human Emotion Using Wavelet Transform and t-ROIs for Fusion of Visible Images and Thermal IR Image Sequences

The visible image-based approach has long been considered the most powerful approach to facial emotion estimation. However it is illumination dependency. Under uncontrolled operating conditions, estimation accuracy degrades significantly. In this paper, we focus on integrating visible images with thermal IR image sequences for facial emotion estimation. First, to address limitations of thermal IR images, such as being opaque to eyeglasses, we apply thermal Regions of Interest (t-ROIs) to sequences of thermal IR images. Then, wavelet transform is applied to visible images. Second, features are selected and fused from visible features and thermal IR features. Third, fusion decision using conventional methods, Principal Component Analysis (PCA) and Eigen-space Method based on class-features (EMC), and our proposed methods, thermal Principal Component Analysis (t-PCA) and norm Eigen-space Method based on class-features (n-EMC), is applied. Applying our suggested methods, experiments on the Kotani Thermal Facial Emotion (KTFE) database show significant improvement, proving its effectiveness.

6.1 Introduction

The detection and estimation of human emotions is a challenging task. In the last decade, automated estimation of human emotions has attracted the interest of many researchers, because such systems will have numerous applications in security, medicine, and especially human-computer interaction. Many previous works [96], [97], [119] proposed have been inclined towards developing facial expression estimation. Nevertheless, there is a lack of accurate and robust facial expression estimation methods to be deployed in uncontrolled environments. When the lighting is dim or when it does not uniformly illuminate the face, the accuracy decreases considerably. Moreover, human emotions estimation based on only the visible spectrum has proved to be difficult in cases where there are emotion changes that expressions do not show. Using thermal IR image, which is not sensitive to light conditions, is a new and innovative way to fill the gap in the human emotions estimation field. Besides, human emotions could be manifested by changing temperature of face skin which is obtained by an IR camera. Consequently, thermal IR image gives us more information to help us robustly estimate human emotions. Although there are many significant advantages when we use IR imagery, it has several drawbacks. Firstly, thermal IR data are subjected to change together with body temperature caused by variable ambient temperatures. Secondly, presence of eyeglasses may result in loss of useful information around the eyes. Glass is opaque to IR, and object made of glass act as temperature screen, completely occluding the parts located behind them. Hence, the sensitivity of IR imagery is decreased by facial occlusions. Thirdly, there are some facial regions not receptive to the emotion changes. To eliminate the effects of these challenging problems above, we propose fusion of visible images and sequences of thermal IR images. To estimate seven emotions, we use the fusion of conventional methods, PCA and EMC, and our proposed methods, t-PCA and n-EMC, over obtained the fusion features.

6.2 Related Work

In the recent years, a number of studies have demonstrated that thermal IR image offers a promising alternative to visible imagery in facial emotion estimation problems by better handling the visible illumination changes. Jarlier et al. [89] extracted the features as representative temperature maps of nine action units (AUs) and used K-nearest neighbor to classify seven expressions. The database for testing has four persons and the accuracy rate is 56.4%. Khan et al. [23] suggested using Facial Thermal Feature Points (FTFPs), which are defined as facial points that undergo significant thermal IR changes in presenting an expression, and used Linear Discriminant Analysis (LDA) to classify intentional facial expressions based on Thermal Intensity Values (TIVs) recorded at the Facial Thermal Feature Points (FTFPs). The database has sixteen persons with five expressions and the accuracy rate ranges from 66.3% to 83.8%. Trujillo et al. [24] proposed using a local and global automatic feature localization procedure to perform facial expression in thermal IR images. They used PCA to reduce the dimension and interest point clustering to estimate facial feature localization and Support Vector Machine (SVM) to classify three expressions. B.Hernández et al. [25] used SVM to classify the expressions "surprise", "happy", "neutral" from two inputs. The first input consists of selections of a set of suitable regions where the feature extraction is performed, second input is the Gray Level Co-occurrence Matrix used to compute region descriptors of the IR images. Nhan et al. [26] extracted time, frequency and time-frequency features from thermal IR data to classify the natural responses in terms of subject-indicated levels of arousal and valence stimulated by the International Affective Picture System. Yoshitomi et al. [33] used two dimensional detection of temperature distribution on the face using infrared rays. Based on studies in the field of psychology, several blocks on the face are chosen for measuring the local temperature difference. With Back Propagation Neutral Network, the facial expression is recognized. The recognition accuracy reaches 90% with "neutral", "happy", "surprising" and "sad" expressions. However, the testing database is obtained from only one female frontal view. Yoshimomi generated feature vectors by using a two-dimensional Discrete Cosine Transformation (2D-DCT) to transform the grayscale values of each block in the facial area of an image into their frequency components, and used them to recognize five expressions, including "angry", "happy", "neutral", "sad", and "surprise". The mean expression accuracy is 80% with four test subjects [90]. Koda et al. used the idea from [90] and added a proposed method for efficiently updating of training data, by only updating the training data with "happy" and "neutral" facial expression after an interval [91]. The expression accuracy increased from 80% to 87% with this new approach. All these studies with thermal IR image have shown that the facial temperature changing is useful for estimating the human emotions.

Recently, a little attention has been paid to facial emotion estimation by using fusion information from visible images and thermal IR information. Wang et al. [32] proposed both decision-level and feature-level fusion methods using visible and IR imagery. In feature-level, they used tools for the Active Appearance Model (AAM) to extract features and extracted three features of head motion for visible feature and calculated several statistical parameters including mean, standard deviation, minimum and maximum as IR features. To select the feature, they used F-test statistic. They also used Bayesians networks (BNs) and SVMs to obtain the feature fusion. In decision-level, BNs and SVMs

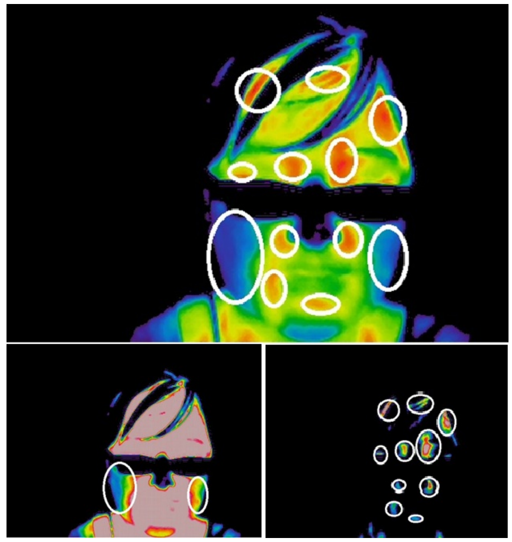


Figure 6.1: An example of t-ROIs.

are used to classify three emotions, happiness, fear and disgust. The results show that their methods improved about 1.35% accuracy compare with only using visible features. Yoshitomi et al. [22] proposed decision-level fusion of voices, visual and IR imagery to recognize the affective states. DCT is used to extract the visible and IR features, then two neutral networks are trained for obtained visible and IR features, respectively. For voice recognition, Hidden Markov Models (HMMs) are used. To decide the results, simple weighted voting is used. Following the related work, there are a few researches using fusion of visible and thermal IR or these approaches that use the extracted features from a single infrared thermal IR image may lose some useful information which could be contained in the sequences. Therefore, we consider two methods of facial emotion estimation by fusing visible images and sequence of thermal IR at decision-level and feature-level respectively.

6.3 Methods

In this section, we propose a feature fusion method to integrate visible images and sequence of thermal IR images by delicate selection of representative features (i.e. t-ROI) in Section 5.1, two classification methods (t-PCA and n-EMC) and a decision-level fusion method which automatically explores the best fusion weights of features in Section 5.2.

6.3.1 Feature-Level Fusion

Before selecting features, we perform some preprocessings such as face normalization, noise deduction.

First, with sequences of thermal IR images, we find the regions of interest based on t-ROIs.

In our definition, interest regions are regions in which temperature increases or decreases significantly when human emotions change. We use the two regions which are the hottest and coldest regions of the face, except the eyeglasses, usually the forehead, eyeholes, and cheek-bone regions, as our interest regions. Before finding the t-ROIs, to avoid any ambient temperature change from frame to frame, we update the temperature of each point of each frame based on the difference between mean of ambient temperature and mean of the first m frame ambient temperature.
Let h be a map from face ($Fa \subset R^2$) to temperature ($T \subset R$) space

$$h : Fa \to T$$

$$(i, j) \mapsto h(i, j)$$

We obtain the t-ROIs by using the following equations:

$$\Delta T_{Fa} = T^{Fa}_{Max} - T^{Fa}_{Min}; \delta T_{Fa} = \Delta T_{Fa}/5$$

$$L^{Fa}_{k,idx} = \{(i, j) \in Fa | T^{Fa}_{Min} + \delta T_{Fa} * (idx - 1) \leq h(i, j) < T^{Fa}_{Max} - \delta T_{Fa} * (5 - idx)\} \quad (6.1)$$

$$L_i dx = L^{Fa}_{k,idx}, k \in \overline{1, K}$$

where $T^{Fa}_{Max}, T^{Fa}_{Min}$ are maximum and minimum of temperature of each human face at frame k, respectively; K is the number of frame; $idx \in \{2, 5\}$.

To estimate the true emotion, using single frame can not give the exact emotion, hence we use the sequence of thermal IR frames. Therefore, after reducing the effect of eyeglasses, to apply for sequence of thermal IR frames, we calculate the accumulate of the discriminant of frame m and frame m + idx.

$$L_i dx : F^* \to T$$

$$(i, j) \mapsto f^*(i, j)$$

We calculate $\sum_m \parallel f^*_m(i, j) - f^*_{m+idx}(i, j) \parallel, \forall (i, j) \in L_i dx$

After obtaining the t-ROIs for each frame, we find the dominant levels of frames. A frame a is more dominant than frame b if only if temperature change of all t-ROIs between frame a and frame a-k is bigger than temperature change of all t-ROIs between frame b and frame b-k. Based on the obtained dominant level of each frame, we can automatically put the weight for each frame.

Let call F is the number of frames. $\forall f_m \in F$

$$w_{f_m} > w_{f_n} \Longleftrightarrow \sum_{\forall i,j \in L_i dx)} \parallel f^*_m(i, j) - f^*_{m+idx}(i, j) \parallel \geq \sum_{\forall i,j \in L_i dx)} \parallel f^*_n(i, j) - f^*_{n+idx}(i, j) \parallel,$$

Second, with visible images, to eliminate of effects of non-uniform illumination and to omit unnecessary details, we use multiresolution analysis with Antonini filter bank. After two level of analysis with 7/9 Antonini filter bank, we keep coeffcients of LL part wavelet transform [116]. Figure 6.2 shows wavelet decomposition at level 1 and 2 over our visible data.

To select the feature between visible feature and thermal IR feature, we perform feature-level fusion of visible and thermal IR image by using t-ROIs and PCA.

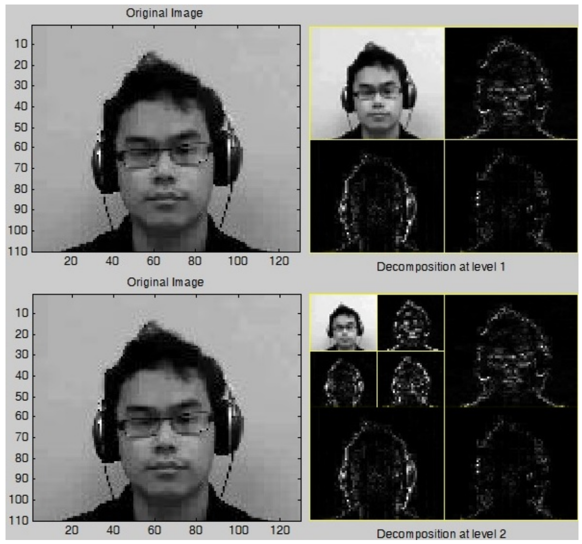


Figure 6.2: Wavelet decomposition at level 1 and 2.

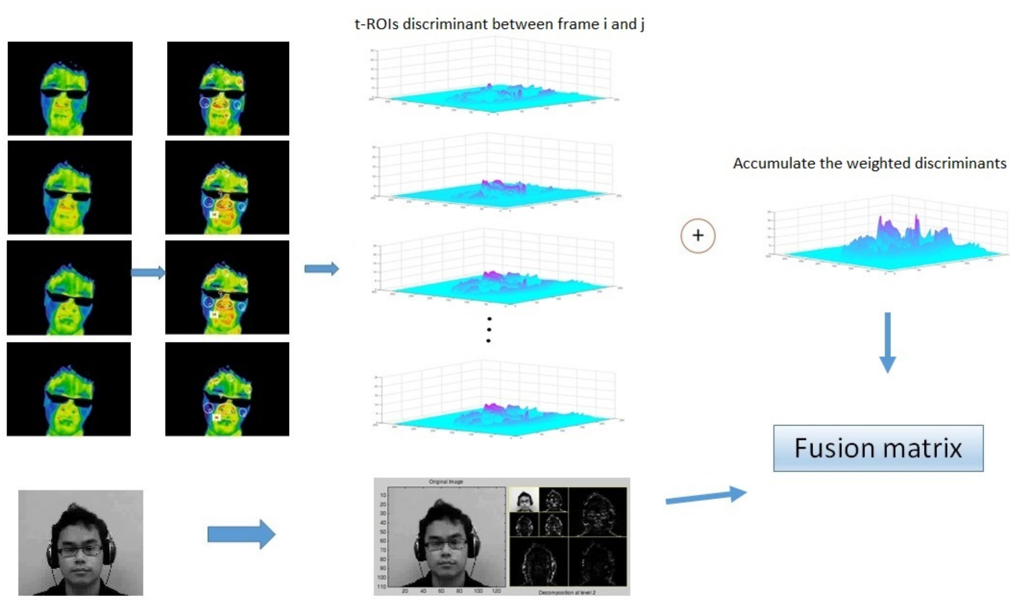


Figure 6.3: A example procedure for fusion of visible images and sequences of thermal IR images.

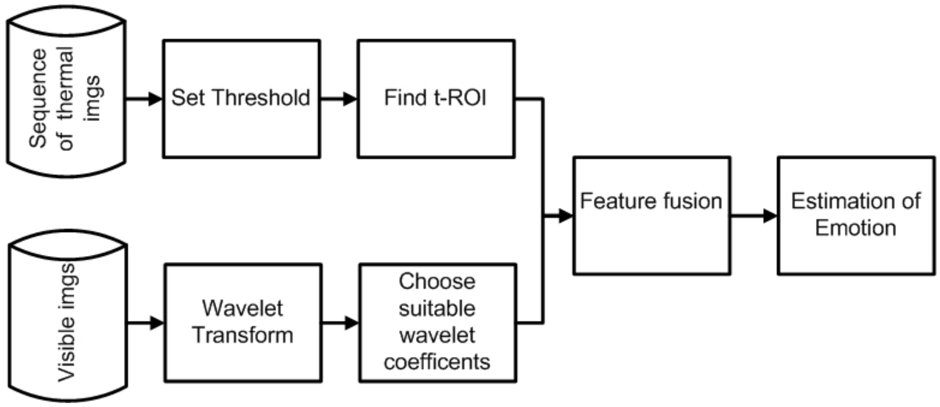


Figure 6.4: Feature fusion of visible and sequence thermal IR image.

Step 1. Find t-ROIs over sequence of thermal IR images.
Step 2. Apply Wavelet transform over visible facial images and keep LL.
Step 3. Apply PCA over accumulate the weighted discriminant t-ROIs.
Step 4. Build matrix from feature vectors obtained from step 2 and 3.
Step 5. Using PCA, EMC, t-PCA and n-EMC to classify emotions.

Fig.6.3 shows a example procedure for fusion of visible images and sequences of thermal IR images.

6.3.2 Decision-Level Fusion

To estimate human emotions, we use a decision fusion method of two conventional methods (PCA, EMC) and our proposed methods, thermal-Principal Component Analysis (t-PCA) and norm-Eigenspace method based on class feature (n-EMC).

With PCA, the aim is to build a face space, including the basis vectors called principal components, which better describes the face images [113]. The difference between PCA and EMC is that PCA finds the eigenvector to maximize the total variance of the projection to line, while EMC is obtained eigenvector to maximize the difference between the within-class and between-class variance [115].

Figure 6.5 shows the procedure of estimating human emotions using t-PCA. Figure 6.6 shows the procedure of estimating human emotions using n-EMC.

The preprocessing for thermal IR data is to normalize data to facial space containing only face. To analyze the human emotion using thermal IR data of emotion, we propose two methods, thermal-Principal Component Analysis (t-PCA) and norm-Eigenspace method based on class feature (n-EMC).

To illustrate the feasibility of using eigen-space to fulfill facial emotion estimation task, thermal PCA (t-PCA) is modified from the PCA reconstruction method and evaluated over thermal IR data [113]. With PCA, the aim is to build a face space, including the basis vectors called principal components, which better describe the face images. PCA has several advantages over other face recognition schemes in its speed and simplicity [113]. We modified PCA to estimate facial emotion from thermal IR data.

Let F be a set of classes to be analyzed. Here, F is a set of all emotion classes. Assume that M_m^f thermal IR frames of training data are given as the facial temperature pattern for each class $f \in F$ where $F = \{anger, disgust, fear, happiness, neutral, sadness, surprise\}$. Let Γ_i^f be the $i-th$ facial temperature pattern where $i = \overline{1, M_f}$; the dimension of Γ_i^f, $n \times m$, is equal to the number of pixels in a thermal IR frame, and each element of Γ_i^f indicates the temperature of each pixel.

Compute the mean of training data $\Psi^f = \frac{1}{M^f}\sum_{i=1}^{M_f}\Gamma_i^f$ and let the normalized vector be $\Phi_i^f = \Gamma_i^f - \Psi^f$. We seek a set of M orthonormal vectors, u_k^f, which best describes the distribution of the training data. The kth vector, u_k^f is chosen by

$$\lambda_k^f = \frac{1}{M_f}\sum_{i=1}^{M_f}((u_k^f)^\tau \Phi_i^f)^2. \tag{6.2}$$

is a maximum, subject to $(u_l^f)^\tau u_k^f = \begin{cases} 1, & \text{if } l = k. \\ 0, & \text{if } otherwise. \end{cases}$

The eigenvectors and eigenvalues are the vector u_k^f and scalar λ_k^f of covariance matrix $C^f = \frac{1}{M_f}\sum_{i=1}^{M_f}(\Phi_i^f(\Phi_i^f)^\tau = A^f(A^f)^\tau$ where $A^f = [\Phi_1^f, \Phi_2^f, ..., \Phi_{M_f}^f]$

$$\Leftrightarrow \lambda_k^f(u_k^f) = C^f u_k^f \tag{6.3}$$

The size of covariance matrix C^f, $nm \times nm$, is too large to determine $n \times m$ eigenvectors and eigenvalues. Turk et al. [118] suggested a computationally feasible method to find these eigenvectors. Let a matrix $H^f = (A^f)^\tau A^f$, then size of H^f is size $M \times M \ll nm \times nm$; let ν_i^f denote the eigenvectors of H^f.

We have $(A^f)^\tau A^f \nu_i^f = \mu_i^f \nu_i^f \Leftrightarrow A^f(A^f)^\tau A^f \nu_i^f = A^f \mu_i^f \nu_i^f \Leftrightarrow A^f(A^f)^\tau A^f \nu_i^f = A^f \mu_i^f \nu_i^f$

$$\Leftrightarrow C^f A^f \nu_i^f = \mu_i^f A^f \nu_i^f \tag{6.4}$$

From Equation(4), $A^f \nu_i^f$ is the eigenvector of $C^f = A^f(A^f)^\tau$ [118]. Therefore we can obtain ρ^f eigenvector of C^f by calculating the $\rho^f(\rho^f \ll nm)$ eigenvectors (ν_i^f) of matrix H^f and multiplying A^f to ν_i^f. After obtaining eigenfaces from training data of each emotion, we map facial thermal IR training data to feature spaces by $\omega_i^f(train) = (u_i^f)^\tau(\Gamma^f - \Psi^f), i = \overline{1, \rho^f}$.

We use the idea that if the input frame is much similar to some emotion training set, the reconstructed data will has less distortion than the data reconstructed from other eigenvectors of training emotions [113]. For each testing facial thermal IR frame Γ_{test}, firstly we project it onto the eigenfaces of each class.

$\omega_{test}^f = (U^f)^\tau(\Gamma_{test} - \Psi^f)$, where $U^f = (u_i^f), i = \overline{1, \rho^f}$.

Secondly, for each emotion, we find the feature which is most similar to the testing projected vector by calculate the angle between vector of training feature space and testing projected vector.

$$\beta^f = \underset{i}{\operatorname{argmax}} \frac{\omega_{test}^f \omega_i^f(train)}{\|\omega_{test}^f\|\|\omega_i^f(train)\|}; i = \overline{1, \rho^f}. \tag{6.5}$$

Thirdly, we find reconstruction of the testing data by the obtained feature in each class.

$\Gamma_{reconst}^f = U^f \omega_{\beta^f}^f + \Psi^f$

Finally, we choose an emotion of which reconstruction of the testing data is the most similarity to testing data.

$$\gamma = \underset{f}{\operatorname{argmax}} \frac{\Gamma_{test}\Gamma^{f}_{reconst}}{\|\Gamma_{test}\|\|\Gamma^{f}_{reconst}\|}; f = \overline{1,7} \tag{6.6}$$

The second valuation to estimate human emotion uses n-EMC over thermal IR data. n-EMC is modified from EMC [114]. The difference between EMC and n-EMC is formulation to calculate the difference between the within-class and between-class variance.

In mathematics, with n-EMC, instead of finding the eigenvectors, u^f_k and eigenvalues λ^f_k of covariance matrix $C^f = \frac{1}{M_f}\sum_{i=1}^{M_f}(\Phi^f_i(\Phi^f_i)^\tau = A^f(A^f)^\tau$ where $A^f = [\Phi^f_1, \Phi^f_2, ..., \Phi^f_{M_f}]$, we find eigenvectors, u^f_k and eigenvalues λ^f_k of matrix $S = \|S_B - S_W\|_2$ where

$$M = \sum_{f \in F} M_f \tag{6.7}$$

$$\Psi^f = \frac{1}{M^f}\sum_{i=1}^{M_f}\Gamma^f_i; \Psi = \frac{1}{M}\sum_{f \in F}\sum_{i=1}^{M_f}\Gamma^f_i \tag{6.8}$$

$$S_B = \frac{1}{M}\sum_{f \in F} M_f\|\Psi_f - \Psi\|_2\|\Psi_f - \Psi\|^\tau_2. \tag{6.9}$$

$$S_W = \frac{1}{M}\sum_{f \in F}\sum_{i=1}^{M_f}\|\Gamma^f_i - \Psi_f\|_2\|\Gamma^f_i - \Psi_f\|^\tau_2. \tag{6.10}$$

For each testing facial thermal IR frame Γ_{test}, firstly we project it onto the eigenfaces of each class.

$\omega^f_{test} = (U^f)^\tau(\Gamma_{test} - \Psi^f)$, where $U^f = (u^f_i), i = \overline{1, \rho^f}$.

Secondly, for each emotion, we find the feature which is most similar to the testing projected vector by calculate the angle between vector of training feature space and testing projected vector.

$$\beta^f = \underset{i}{\operatorname{argmax}} \frac{\omega^f_{test}\omega^f_i(train)}{\|\omega^f_{test}\|\|\omega^f_i(train)\|}; i = \overline{1, \rho^f}. \tag{6.11}$$

Finally, we choose an emotion which has maximum of β^f

$$\gamma = \underset{f}{\operatorname{argmax}}\, \beta^f; f = \overline{1,7} \tag{6.12}$$

To estimate human emotions, we use decision fusion method of PCA, t-PCA, EMC and n-EMC. Figure 6.7 shows the general procedure to estimate human emotions using decision fusion.

When using decision fusion of PCA (t-PCA), we used the estimation of emotion module as described in figure 6.5. To determine the best class of emotions, after using PCA (t-PCA), the voting method with weights is used. The weights are set to fusion data and visible image, respectively. We determine the emotion class f of input image by choosing j satisfied minimum of following equation:

$$f = argmin\left(w_1 * MSE^{VI}_j + w_2 * MSE^{FU}_j\right) \tag{6.13}$$

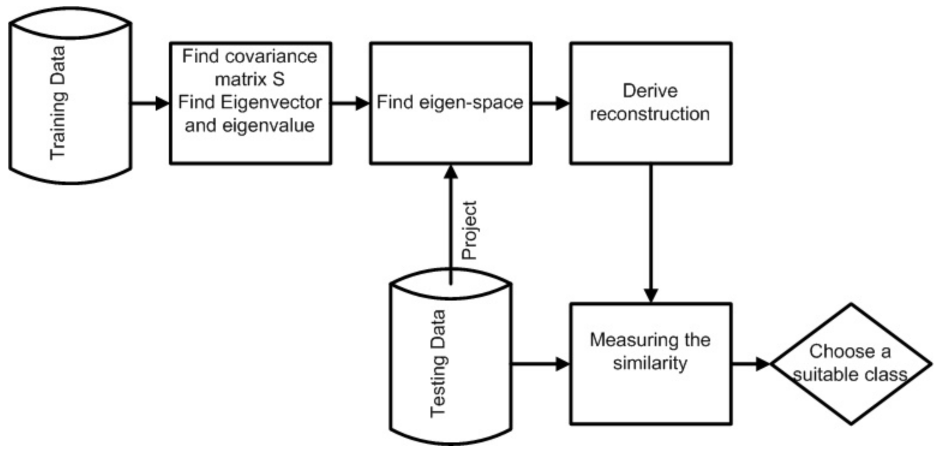


Figure 6.5: Estimation of emotion using t-PCA.

where MSE_j^{VI} and MSE_j^{FU} are mean square errors calculated at class j of visible image and fusion data. We set $w_1 = \frac{4}{3}$ and $w_2 = \frac{2}{3}$ in experiment.

To estimate human emotion using decision fusion of EMC (n-EMC), we used the estimation of emotion module as described in figure 6.6. To determine the best class of emotions, after using EMC (n-EMC), the voting method with weights is used. The weights are set to fusion data and visible image, respectively.

We determine the emotion class f of input image by choosing j satisfied maximum of following equation:

$$k = max_i \frac{f_h^{VI} * F_i^{VI}}{\|f_h^{VI}\| * \|F_i^{VI}\|}, i = \overline{1,n} \tag{6.14}$$

$$g = max_i \frac{f_h^{FU} * F_i^{FU}}{\|f_h^{FU}\| * \|F_i^{FU}\|}, i = \overline{1,n} \tag{6.15}$$

$$f = argmax\left(w_1 * k + w_2 * g\right), \tag{6.16}$$

where n is a number of the training images of class j; f_h^{VI} and f_h^{FU} are testing image h of visible image and fusion data, respectively; F_i^{VI} and F_i^{FU} are vector i of eigenface of visible image and fusion data, respectively. We set $w_1 = \frac{2}{3}$ and $w_2 = \frac{4}{3}$ in experiment.

6.4 Database

The KTFE database [112] includes 186.2GB visible and thermal facial emotion videos, visible facial expression image database and thermal facial expression image database. This database contains 30 subjects who are Vietnamese, Japanese, Thai and Chinese from 11 year-old to 32 year-old with seven emotions.

From draw data of KTFE database, we extract manually visible images and sequences of thermal IR images based on self-reports of participants, expressions and changing of facial temperatures. Causing the time-lag phenomenon, the sequence of thermal IR images are designed from a frame which we extracted the visible image to a frame which

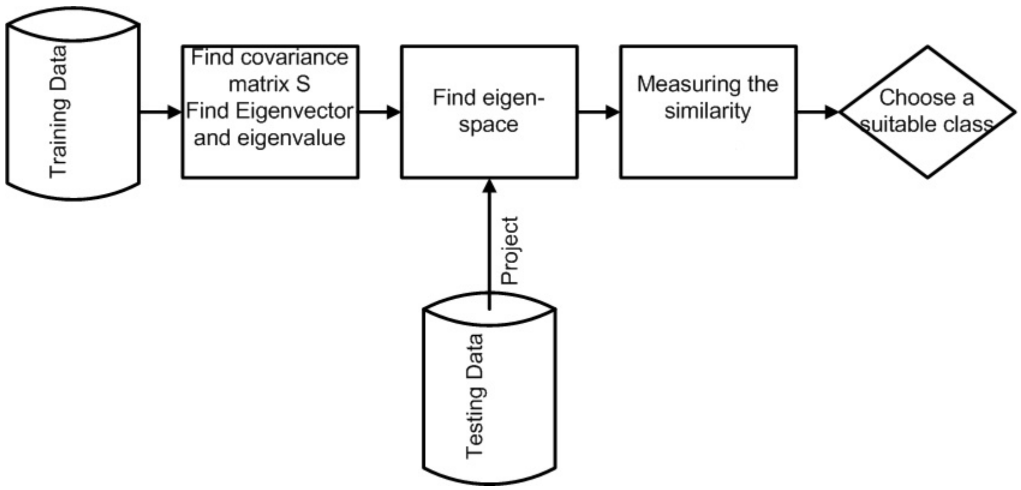


Figure 6.6: Estimation of emotion using n-EMC.

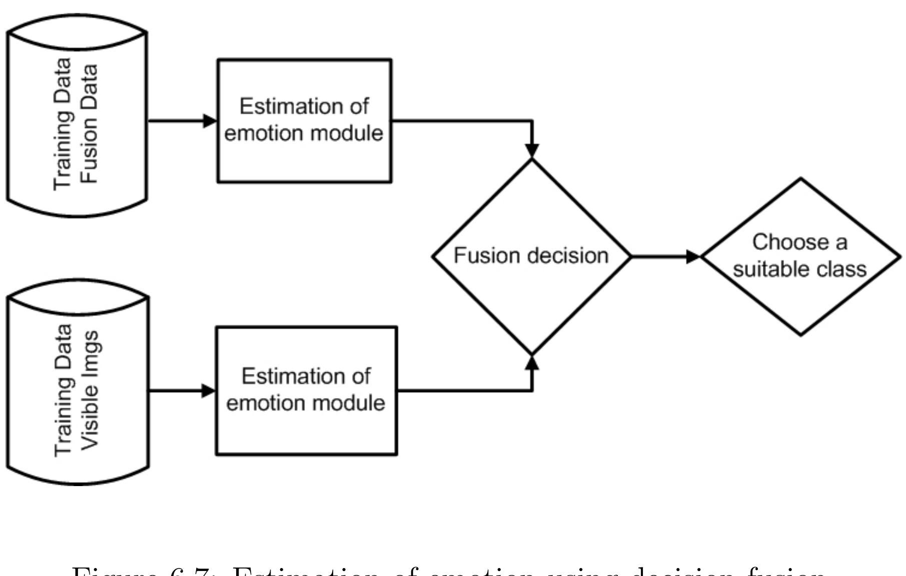


Figure 6.7: Estimation of emotion using decision fusion.

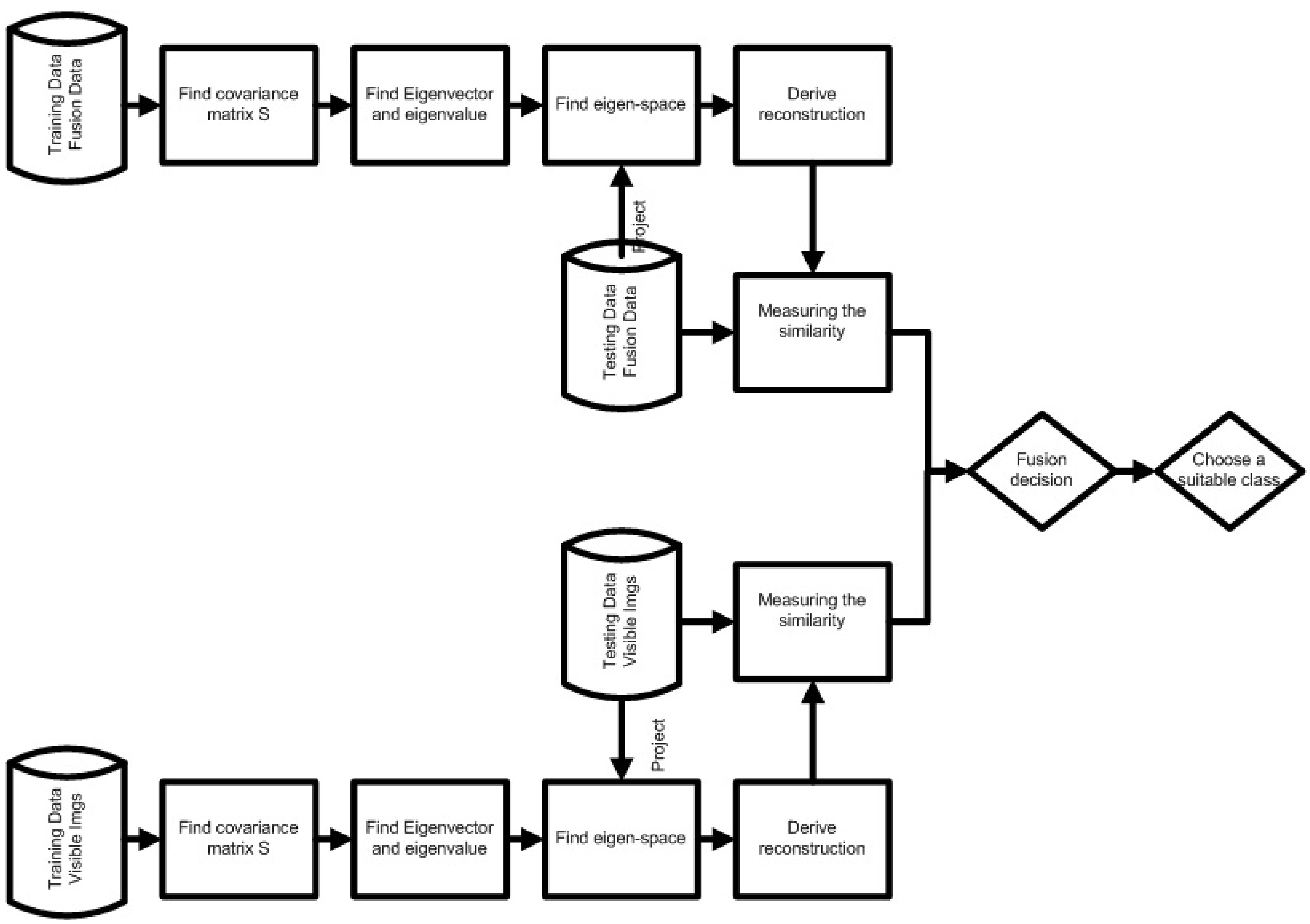


Figure 6.8: Estimation of emotion using t-PCA fusion.

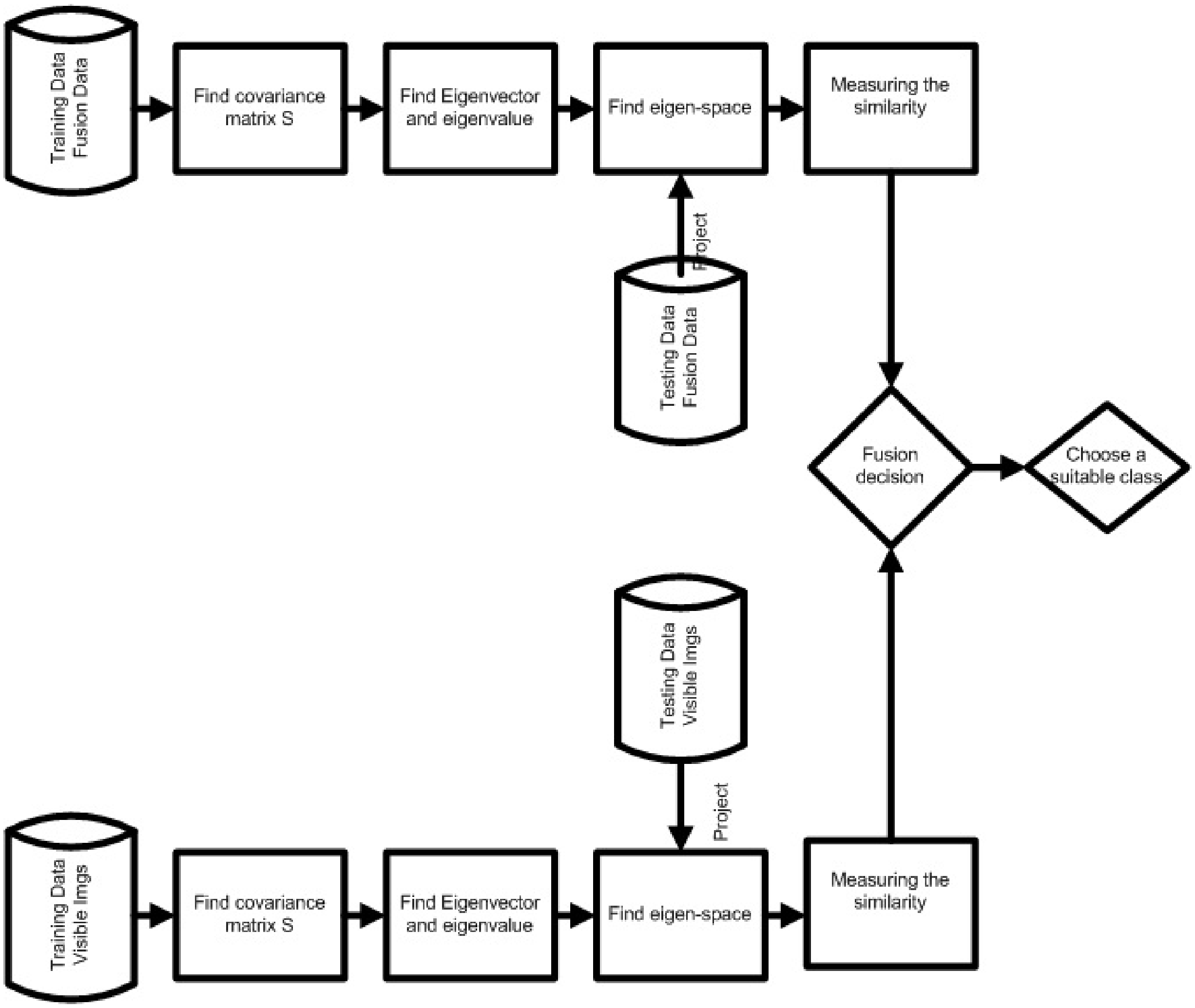


Figure 6.9: Estimation of emotion using n-EMC fusion.

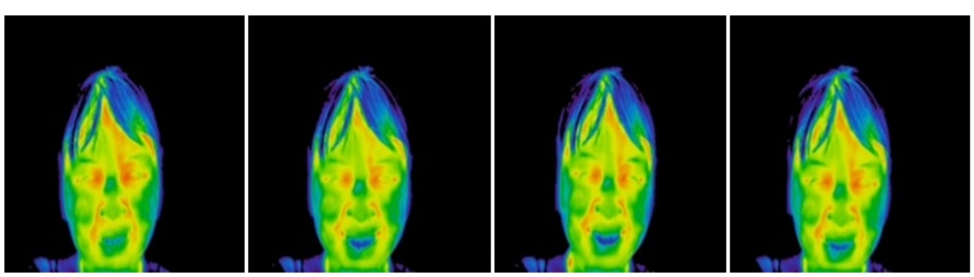


Figure 6.10: Sample sequence of thermal images.

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	89.16	2.80	-	2.79	-	5.25	-
Disgust	1.26	84.90	-	-	-	-	13.84
Fear	1.32	-	83.51	8.37	1.32	5.48	-
Happiness	-	-	-	90.81	-	9.19	-
Neutral	12.21	-	-	-	86.23	1.55	-
Sadness	-	-	14.60	1.42	-	83.98	-
Surprise	-	8.28	-	-	-	8.28	83.44
Average							**86.00**

Table 6.1: The confusion matrix of EMC with fusion of visible image and thermal IR image sequence

is after the participant emotion is neutral. Fig.6.10 shows a sample sequence of thermal IR images.

From raw data of KTFE database, we extract manually visible images and sequences of thermal IR images based on self-reports of participants, expressions and changing of facial temperatures. Causing the time-lag phenomenon, the sequence of thermal IR images are designed from a frame which we extracted the visible image to a frame which is after the participant emotion is neutral.

6.5 Experimental Results

In our experiments, we separate the training and testing data as 60% and 40% of total visible images, sequence of thermal IR images, and fusion of visible image and thermal IR image sequence.

Fig.6.11 shows the results of facial emotion estimation of EMC with visible images (vi_EMC), sequence of thermal IR images (ther_EMC) and fusion of visible images and sequences of thermal IR images (fu_EMC). Following [117], accuracy of estimating human emotion using thermal IR images is lower than using visible images. Because emotions of thermal IR images are always not clearer than emotions of visible images. Therefore, with EMC methods, good for classification, the results using visible images are better than results using thermal IR images. However, with our new results, accuracy of estimating human emotion using sequences of thermal IR images is better than using visible images. When using fusion information, the average accuracy increases 2.77% in compare with using only visible features, especially for happiness. In general, average accuracy of each emotion increases when we use fusion information. The results prove the necessary of fusion information.

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	90.80	2.73	-	2.72	-	3.75	-
Disgust	3.16	85.32	-	-	-	-	11.52
Fear	2.35	-	84.39	4.62	1.11	7.53	-
Happiness	-	-	-	90.96	-	9.04	-
Neutral	8.57	-	-	-	87.14	4.29	-
Sadness	-	-	12.16	3.16	-	84.68	-
Surprise	-	7.75	-	-	-	7.75	84.50
Average							**86.83**

Table 6.2: The confusion matrix of n-EMC with fusion of visible image and thermal IR image sequence

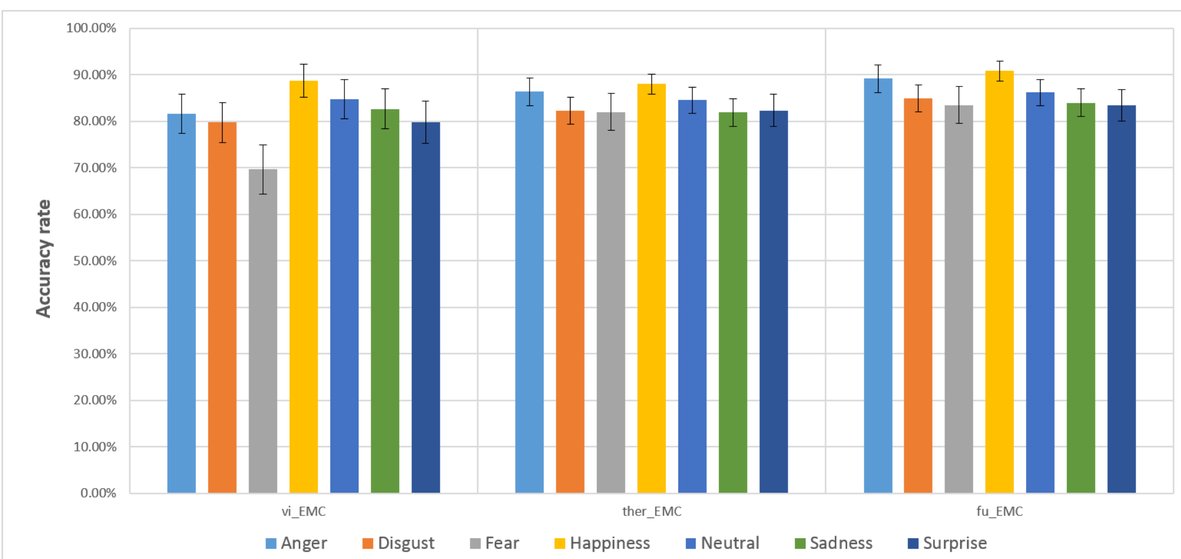


Figure 6.11: The emotion estimation results of ECM with fusion of visible image and thermal IR image sequence

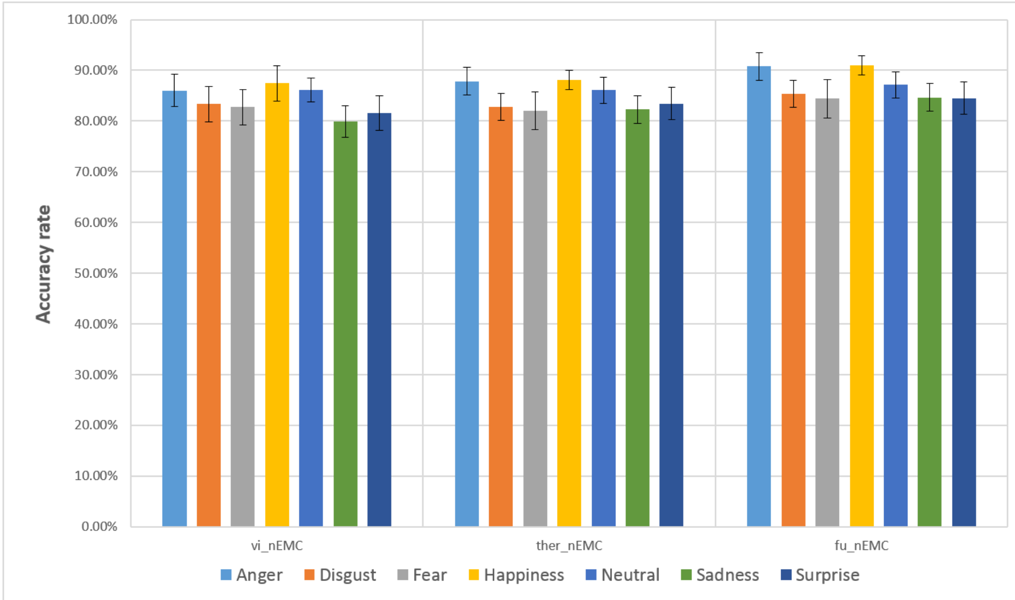


Figure 6.12: The emotion estimation results of n-EMC with fusion of visible image and thermal IR image sequence

Fig.6.12 shows the results of facial emotion estimation of n-EMC with visible images (vi_n-EMC), sequence of thermal IR images (ther_n-EMC) and fusion of visible images and sequences of thermal IR images (fu_n-EMC). With n-EMC, the average accuracy using visible images, sequences of thermal IR images and fusion of visible images and sequences of thermal IR images increases 3.15%, 2.41%, 2.35%, respectively compared with EMC. Similar to the results using EMC, the accuracy using fusion data is better than using other data.

Fig.6.13 shows the results of facial emotion estimation of PCA with visible images (vi_PCA), sequences of thermal IR images (ther_PCA) and fusion of visible images and sequences of thermal IR images (fu_PCA). With PCA, accuracy using sequence of thermal IR images is better than accuracy using visible images. PCA works worse than EMC, which is good to classify each emotion. In general, with PCA, using fusion data gives the best results comparing using thermal IR images, visible images and sequence of thermal IR images.

Fig.6.14 shows the results of facial emotion estimation of t-PCA with visible images (vi_t-PCA), sequences of thermal IR images (ther_t-PCA) and fusion of visible images and sequences of thermal IR images (fu_t-PCA). With t-PCA, the average accuracy using visible images, sequences of thermal IR images and fusion of visible images and sequences of thermal IR images increases 2.03%, 1.31%, 0.48% respectively in compare with PCA. Especially, for disgust with visible images, accuracy improvement is 11.97%. Our method, t-PCA, yields an average improvement of 2.33% in performance of facial emotion estimation compared with using only visible features.

In conclusion, comparing the results of visible images, sequences of thermal IR images, and fusion data, the accuracy of estimating emotion using fusion data is better than the accuracy of estimation emotion using only visible images, thermal IR images and sequences

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	87.16	1.36	3.88	2.86	-	1.36	3.38
Disgust	-	85.03	2.31	0.63	-	2.00	10.03
Fear	3.91	-	85.04	4.70	1.76	4.59	-
Happiness	-	-	-	84.32	-	15.68	-
Neutral	5.10	-	1.88	3.43	85.83	3.76	-
Sadness	1.05	-	2.82	4.77	3.03	88.33	-
Surprise	3.45	12.28	-	2.00	-	-	82.27
Average							**85.43**

Table 6.3: The confusion matrix of PCA with fusion of visible image and thermal IR image sequence

of thermal IR images.

6.6 Conclusions

In this chapter, we have proposed the fusion of visible features and thermal IR features for estimating human emotions. Specially, two classification methods, t-PCA and n-EMC, are proposed to perform decision-level fusion. Our experiment on KTFE spontaneous database show that our methods yield and average improvement of 2.58% in performance of facial emotion estimation compared with using only visible features. Our method has several advantaged points. First, to the best of our knowledge, this is one of the first methods using sequences of thermal IR images. Emotion is complex action of human. To understand it clearly, using a single image cannot figure out the exact emotion. Besides, using thermal IR information with single frame cannot give the right emotion. Therefore, it is necessary to use sequences of thermal IR images. Second, with t-ROIs, we fill the gaps of thermal IR image, eyeglass problem. Third, using the weight discriminant features help our method decrease the running cost and increase accuracy rate. Because there are some frames more important than others, the weights set for frames are necessary. Fourth, using wavelet transform for visible image gives several advantages such as to reduce the unnecessary coarse, and so on. The fusion features, obtained from important visible features and necessary thermal IR feature, are better than only visible and thermal IR features. Two proposed classification methods, t-PCA and n-EMC, are successful to improve estimation accuracy. We also suggest decision fusion with weighted similarity measure for the conventional methods (PCA and EMC) and the our proposed methods (t-PCA, n-EMC) to increase the estimation accuracy. Experiments are tested in fusion database, specially designed from KTFE database. The results prove that the fusion of

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	87.79	1.36	1.33	3.50	1.91	1.36	2.75
Disgust	-	88.91	2.31	-	-	-	8.78
Fear	-	0.42	86.54	5.51	3.19	3.51	0.83
Happiness	-	0.56	0.91	90.08	0.45	8.00	-
Neutral	1.67	-	3.02	2.29	87.59	5.43	-
Sadness	-	1.67	2.12	4.35	0.53	91.33	-
Surprise	3.45	12.61	-	-	-	-	84.94
Average							**88.17**

Table 6.4: The confusion matrix of t-PCA with fusion of visible image and thermal IR image sequence

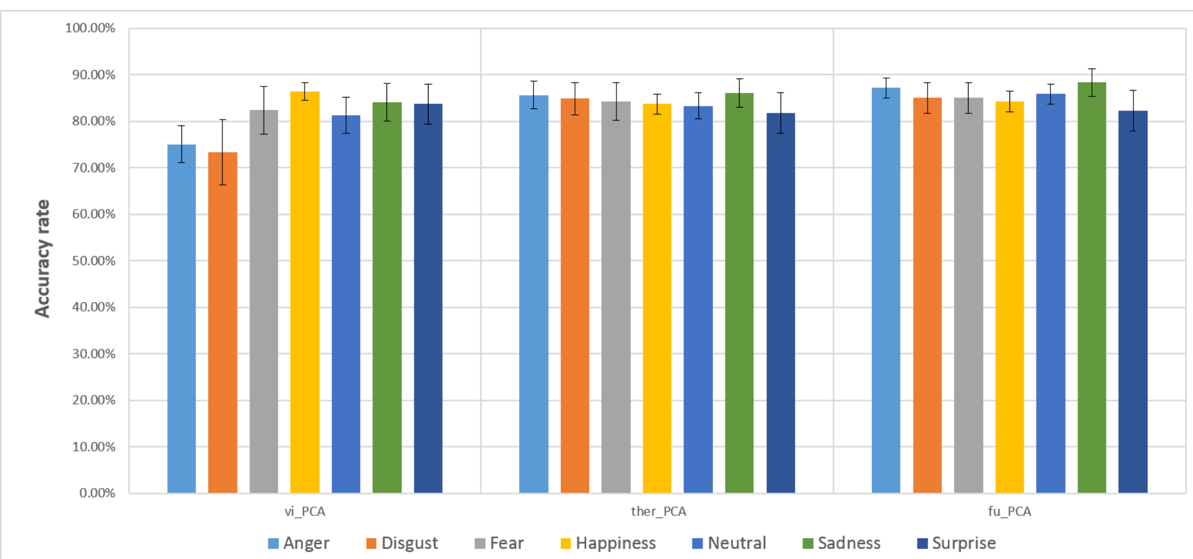


Figure 6.13: The emotion estimation results of PCA with fusion of visible image and thermal IR image sequence

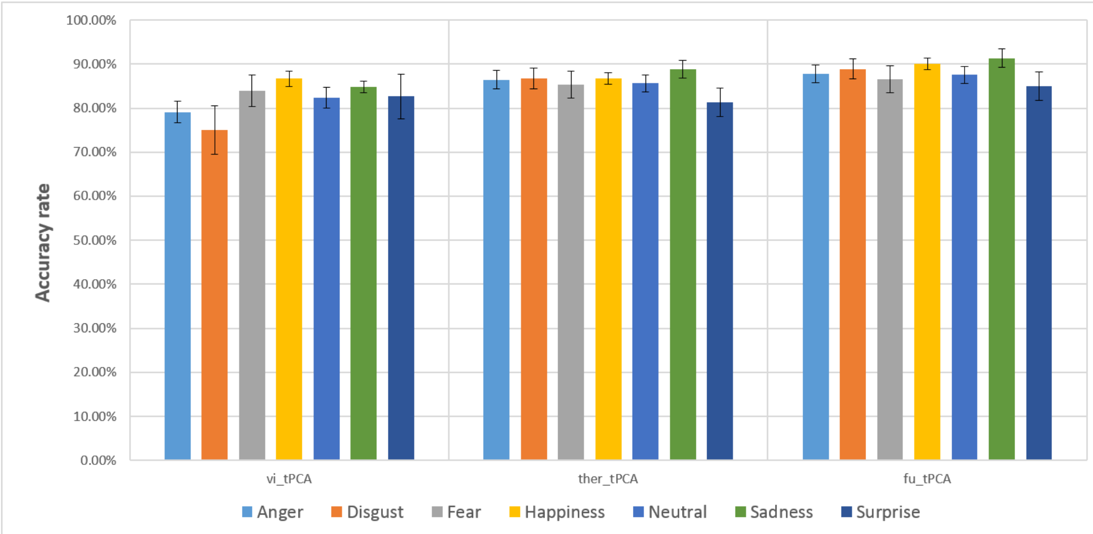


Figure 6.14: The emotion estimation results of t-PCA with fusion of visible image and thermal IR image sequence

visible images and thermal IR image sequences performs better than either of the data. As a future work, we plant to improve the t-ROIs considerably, and also investigate more sophisticated fusion techniques for visible images and thermal IR image sequences and sequences of visible and thermal IR images.

Chapter 7

Conclusions and Future Works

7.1 Major Contribution

In this thesis, we focus on solving the several problems in the automatic facial emotion estimation system. To estimate the true emotion, the problems we focus on are: the effect of eyeglasses on IR data, multi-modal visible and thermal facial emotion database, thermal regions of interest, decision level, thermal image sequence, and fusion of visible images and thermal image sequence. We summarize our main contribution on solving these problems:

1. A multi-modal visible and thermal facial emotion database is proposed and established. With careful and precise procedures from participant preparation, environment preparation, data acquisition, the KTFE database has several advantages:

    - Firstly, to the best of our knowledge, this is one of the first natural spontaneous visible and thermal videos. These databases will allow researchers on facial expressions and emotions to have more approaches more realistic;
    - Secondly, this database already fixed some mistakes which the former database made met when they did experiment settings such as the time-lag phenomenon;
    - Thirdly, we also had several analysis in our data and obtained some interesting results to support research in emotion estimation and human emotion using our database.

2. We also developed a novel way to reduce the effect of changing ambient temperature which occurs during the data acquisition and the proposed method to reduce the effect of eyeglasses using temperature space in thermal data. Since the eyeglasses areas are replaced with the averaged-ambient temperature, these is no more effect of ambient temperature to these areas. The experiment with after removing glasses and before removing glasses shows the increasing of accuracy rate. Specially, with t-PCA, the accuracy of angry emotion increases from 75.63% to 84.38%.

3. In order to handle the eyeglasses problem and focus on main facial regions which receive the emotion changes, we first investigated the influence of generally dominant facial parts, then defined and provided t-ROIs. For the first try t-ROIs, not only the system speed is increased in comparison with the former but also the average of estimation accuracy. For each emotion, the temperature change is different. Therefore, using only static image is not enough to obtain the true emotions. The thermal image sequences are used to obtain the human emotion.

4. We propose the fusion of visible features and thermal features for estimating human emotions. Our method has several advantaged points.

    - First, to the best of our knowledge, this is one of the first methods using sequence of thermal images. Emotions are complex actions of human. To understand it clearly, using a single image is not enough to figure out the exact emotion. Besides, using thermal information with single frame can not give the right emotion. Therefore, it is necessary to use sequence of thermal images.
    - Second, with t-ROIs, we fill the gaps of thermal image, eyeglass problem.
    - Third, using wavelet transform for visible image gives several advantages such as to reduce the unnecessary information, so on. The fusion features, obtained

from important visible features and necessary thermal feature, are better than only visible and thermal features. The results prove that the fusion of visible images and thermal image sequences performs better than either of the data.

  - Finally, the t-PCA and n-EMC are proposed by us to be suitable for decision fusion level.

5. Our research is an epoch-making and challenging research that makes it possible to presume complex feelings other than the famous six basic expressions of Ekman and others.

7.2 Limitations

1. Most of the participants are Vietnamese, just a few Thais, and Japanese. Therefore, it does totally reflect general human emotion of all human being on the earth.

2. In general, the accuracy of each emotion and each method has been improved with the correction of eyeglass areas. However, in some case, the loss of information causes the decrease of accuracy rate.

3. The first version of t-ROIs is still simple and has a problem with heating obstacles.

7.3 Future Works

To improve our database, we intend to obtain more data from participants which have different cultural, education background, nationalities, ages. We also try to acquire more data from real environment such as school environment, supermarket environment, home environment and so on. And if we have more thermal equipment, we intent conduct a research on multi-views for thermal-based data. And with EMG (Electromyogram) and thermal equipment, we will study on the relationship of emotion and musculo-physiological activities.

With our database, we will continue to research more on the relationship between human expressions and human emotions and the relationship between temperature and emotions. From those knowledge, we will build up more real world applications that can support better Human-Computer Interaction.

We intent to study more about t-ROIs and design the good one more suitable with this concept.

From the results of chapter 5, the fusion model for visible-based and thermal-based information promise good results when we conduct more fusion models.

Based on the our research result, it is promising to build up the new complex facial emotion model including facial expression, thermal information and time axes.

Future aspects of the research could be applied in commercial application such as elder caring robot, surveillance tasks and so on.

Last but not least, facial emotion estimation will be an open field for multiple areas such as computer science, behavioral science, psychology and medical science. With all kinds of technologies from those fields, we hope we can achieve major breakthrough and anticipate that in a not far away future, we can enjoy the services from the technologies from facial emotion estimation.

Bibliography

[1] C.Bartneck.: “How convincing is Mr. Datas smile: Affective expressions of machines”, User Modeling and User-Adapted Interaction, vol. 11, pp. 279-295 (2001).

[2] R.A..Brooks.: “Flesh and Machines: How Robots will change us”, Pantheon Books, New York (2002).

[3] R.W.Picard.: “Affective Computing”, MA: MIT Press (2000).

[4] P.Redford.: “Conference Report: Social goals and emotions”, The Psychologist, vol. 13, no. 6, pp. 290-291 (2000).

[5] C.Busso, Z.Deng, S.Yildirim, M.Bulut, C.-M. Lee, A.Kazemzadeh, S.Lee, U.Neumann, S.Narayanan.: “Analysis of emotion recognition using facial expressions, speech and multimodal information”, In the Proceedings of 6th International Conference on Multimodal Interface, ICMI04, PA, pp. 205-211 (2004).

[6] J.Klein, Y.Moon, R.W.Picard.: “This computer responds to user frustration: Theory, design and results”, Interacting with computers, vol. 14, no. 2, pp.119-140 (2002).

[7] I.Cohen, N.Sebe, A.Garg, L.S.Chen, T.S.Huang.: “Facial expression recognition from video sequences: temporal and static modeling”, Journal of Computer Vision and Image Understanding, no. 91, pp. 160-187 (2003).

[8] C.K.Eveland, D.A.Socolinsky, L.B.Wolf.: “Tracking human faces in infrared video”, Image and Vision Computing, vol. 21, no. 7, pp. 579-590 (2003).

[9] D.A.Socolinsky, A.Selinger, J.D.Neuheisel.: “Face recognition with visible and thermal infrared imagery”, Computer Vision and Image Understanding, vol. 91, pp. 72-114 (2003).

[10] C.S.Lisetti and D.J.Schiano.: “Automatic facial expression interpretation: Where human-computer interaction, artificial intelligence and cognitive science intersect”, Pragmatics and cognition, vol. 8, no. 1, pp. 185-235 (2000).

[11] B.Reeves and C.Nass.: “The media equation: How People Treat Computers, Television and New Media Like Real People and Places”, London: Cambridge University Press (1996).

[12] I.Essa and A.Pentland.: “Coding, analysis, interpretation and recognition of facial expressions”, IEEE Transactions on Pattern Analysis, Machine Intelligence, vol. 19, no. 7, pp. 757-763 (1997).

[13] Y.Gao, M.K.H.Leung,S.C.Hui, M.W.Tananda.: "Facial expression recognition form line-based caricatures", IEEE Transactions on Systems, Man, and Cybernetics, vol. 33, no. 3, pp. 407-412 (2003).

[14] S.Morishima.: "Face analysis and synthesis", IEEE Signal Processing Magazine, vol. 18, no. 3, pp. 26-34 (2001).

[15] Y.Du and X.Lin.: "Emotional facial expression model building", Pattern Recognition Letters, vol. 24, pp. 2923-2934 (2003).

[16] W.Y.Andrew.: "Face and mind", Oxford Cognitive Science, Oxford University Express (1998).

[17] M.M.Khan.: "Cluster-analytic classification of facial expressions using infrared measurements of facial thermal features", Doctoral thesis, University of Huddersfield (2008).

[18] T.J.Phillips.: "High performance thermal imaging technology", Advanced Semiconductor Magazine, vol. 15, no. 7, pp. 32-36 (2002).

[19] B.F.Jones and P.Plassmann.: "Digital infrared thermal imaging of human skin", IEEE Engineering in medicine and biology, vol. 21, no.6, pp. 41-48 (2002).

[20] K.Otsuka, S.Okada, M.Hassan, T.Togawa.: "Imaging of skin thermal properties with estimation of ambient radiation", IEEE Engineering in Medicine and Biology, vol. 21, no. 6, pp. 49-55 (2002).

[21] M.Bales.: "High-resolution infrared technology for softtissue injury detection", IEEE Engineering in Medicine and Biology, vol. 17, pp. 56-59 (1998).

[22] Y.Yoshitomi, S-I.Kim, T.Kawano and T.Kitazoe.: "Effects of sensor fusion for recognition of emotional states using voice, face image and thermal image of face", In the Proceedings of the IEEE International workshop on Robotics and Human Interactive Communication, Osaka, Japan, September 2000, pp. 178-183 (2000).

[23] M.M.Khan,R.D.Ward, ,M.Ingleby.: "Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature", ACM Transactions on Applied Perception, vol. 6, pp. 122 (2009).

[24] L.Trujillo, G.Olague, R.Hammoud, B.Hernandez.: "Automatic feature localization in thermal images for facial expression recognition", IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, CVPR Workshops, p. 14 (2005).

[25] B.Hernández, G.Olague, R.Hammoud, L.Trujillo, E. Romero.: "Visual learning of texture descriptors for facial expression recognition in thermal imagery", Computer Vision and Image Understanding, vol. 106, pp. 258269 (2007).

[26] B.R.Nhan and T.Chau.: "Classifying affective states using thermal infrared imaging of the human face", IEEE Transactions on Biomedical Engineering, vol. 57, pp. 979987 (2010).

[27] Z.Liu and S.Wang.: "Emtion recognition using Hidden Markov Model from facial temperature sequence", LNCS 6975, pp.240-247 (2011).

[28] L.V.Yanpen and S.Wang.: "A spontaneous facial expression recognition method using head motion and AAM features", Nature and Biologically Inspired Computing (NaBIC), 2010 Second World Congress, pp.334,339 (2011).

[29] Y.Li, S.Wang, Y.Zhao, Q.Ji.: "Simultaneous Facial Feature Tracking and Facial Expression Recognition", Image Processing, IEEE Transactions, vol.22, no.7, pp.2559-2573 (2013).

[30] P.Shen, S.Wang, Z.Liu.: "Facial Expression Recognition from Infrared Thermal Videos", Intelligent Autonomous Systems 12 LNCS , vol.194, no.7, pp.323-333 (2013).

[31] S.Wang and S.He.: "Spontaneous Facial Expression Recognition by Fusing Thermal Infrared and Visible Images", Intelligent Autonomous Systems 12 LNCS , vol.194, no.7, pp.263-272 (2013).

[32] S.Wang, S.He, Y.Wu, M.He, Q.Ji.: "Fusion of visible and thermal images for facial expression recognition", J.Frontiers of Computer Science , vol.8, no.2, pp.232-242 (2014).

[33] Y.Yoshitomi,N.Miyawaki, S.Tomita, S.Kimura.: "Facial expression recognition using thermal image processing and neural network", Robot and Human Communication, RO-MAN '97. Proceedings., 6th IEEE International Workshop, pp.380-385 (1997).

[34] J.Allanson and S.H.Fairclough.: "A research agenda for physiological computing", Interacting with computers, vol. 16, pp. 857-878 (2004).

[35] V.Bettadapura.: "Face expression recognition and analysis: the state of the art", arXiv preprint arXiv:1203.6722 (2012).

[36] C. Darwin.: "The expression of the emotions in man and animals", New York: Oxford University (1872/1998).

[37] A. Mehrabian.: "Nonverbal communication", Transaction Publishers (1977).

[38] D.M.Sloan, M.M.Bradley,E.Dimoulas, P.J.Lang.: "Looking at facial expressions: Dysphoria and facial EMG", Journal of Biological Psychology, vol. 60, pp. 79-90 (2002).

[39] C.L.Lisetti and F.Nasoz.: "Using non-invasive wearable computers to recognize human emotions from physiological signals", EURASIP Journal of Applied Signal Processing, no. 11, pp. 1672-1687(2004).

[40] H.G.Hosseini and Z.Krechowec.: "Facial expression analysis for estimating patients emotional states in RPMS", In the Proceedings of the 2004 IEEE International Conference on Engineering in Medicine and Biology, IEEE EMBC 2004, vol. 2, pp. 1517-1520 (2004).

[41] G.H. Shergill, H. Sarrafzadeh, O. Diegel, A. Shekar.: "Computerized sales assistants: The application of computer technology to measure consumer interest;a conceptual framework", In the Proceedings of the 2004 IEEE International Conference on Engineering in Medicine and Biology, Journal of Electronic Commerce Research, vol. 9, no. 2, pp. 176191 (2008).

[42] A. van Dam.: "Beyond wimp", Computer Graphics and Applications, vol. 20, no. 1, pp. 5051 (2000).

[43] V.W. Zue and J.R. Glass.: "Conversational interfaces: Advances and challenges", Proceedings of the IEEE, vol. 88, no. 8, pp. 11661180 (2002).

[44] A. Pentland.: "Looking at people: Sensing for ubiquitous and wearable computing", Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 107119 (2000).

[45] E. Vural, M. Bartlett, G. Littlewort, M. Cetin, A. Ercil, J. Movellan.: "Discrimination of moderate and acute drowsiness based on spontaneous facial expressions", In ICPR (2010).

[46] Y. Dai, Y. Shibata, T. Ishii, K. Hashimoto, K. Katamachi, K. Noguchi, N. Kakizaki, D. Ca.: "An associate memory model of facial expressions and its application in facial expression recognition of patients on bed", In ICME, pp. 591 594 (2001).

[47] P. Lucey, J.F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett,K.M. Prkachin.: "Automatically Detecting Pain in Video Through Facial Action Units", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 99, pp. 111 (2010).

[48] E.E.Forbes, J.F.Cohn, N.B. Allen, P.M.Lewinsohn.: "Infant affect during parent-infant interaction at 3 and 6 months: Differences between mothers and fathers and influence of parent history of depression", Infancy, vol. 5, pp. 6184 (2004).

[49] M. Madsen, R. el Kaliouby, M. Eckhardt, M. Hoque, M. Goodwin, R.W. Picard.: "Lessons from participatory design with adolescents on the autism spectrum", In Proc. Computer Human Interaction (2009).

[50] M. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan.: "Real time face detection and facial expression recognition: Development and applications to human computer interaction", In CVPR Workshops for HCI (2003).

[51] V. Bruce.: "What the human face tells the human mind: Some challenges for the robot-human interface", In IEEE Int. Workshop on Robot and Human Communication (1992).

[52] H. Lo and R. Chung.: "Facial expression recognition approach for performance animation", In International Workshop on Digital and Computational Video (2001).

[53] B. J. Theobald and J. F. Cohn.: "Facial image synthesis", Oxford University Press (2009).

[54] D. Huang and F.D.L.Torre.: "Bilinear kernel reduced rank regression for facial expression synthesis", In ECCV (2010).

[55] F.D.L.Torre and J. F. Cohn .: "Facial expression analysis", In Visual Analysis of Humans, Springer, pp. 377-409 (2011).

[56] Y. Tian, T. Kanade, J. Cohn .: "Recognizing Action Units for Facial Expression Analysis", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 97 115 (2001).

[57] F. Bourel, C.C. Chibelushi, A. A. Low .: "Recognition of Facial Expressions in the Presence of Occlusion", Proc. of the Twelfth British Machine Vision Conference, vol. 1, pp. 213 222 (2001).

[58] M. Pardas and A. Bonafonte .: "Facial animation parameters extraction and expression recognition using Hidden Markov Models", Signal Processing: Image Communication, vol. 17, pp. 675 688 (2002).

[59] I. Cohen, N. Sebe, A. Garg, L.S. Chen, T.S. Huang .: "Facial Expression Recognition From Video Sequences: Temporal and Static Modeling", Computer Vision and Image Understanding, vol. 91, pp. 160- 187 (2003).

[60] I. Cohen, N. Sebe, F. Cozman, M. Cirelo, T. Huang .: "Learning Bayesian Network Classifiers for Facial Expression Recognition Using Both Labeled and Unlabeled Data", Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. I-595-I-604 (2003).

[61] P. Michel and R. Kaliouby.: "Real Time Facial Expression Recogni tion in Video Using Support Vector Machines",Proc. 5th Int. Conf. Multimodal Interfaces, Vancouver, BC, Canada, pp. 258 264 (2003).

[62] M. Pantic and J.M. Rothkrantz.: "Facial Action Recognition for Facial Expression Analysis from Static Face Images",IEEE Trans.Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on , vol.34, no.3, pp. 1449 -1461 (2004).

[63] I. Buciu and I. Pitas.: "Application of Non - Negative and Local Non Negative Matrix Factorization to Facial Expression Recognition",Proc. ICPR, pp. 288 291 (2004).

[64] M. Pantic and I. Patras.: "Detecting facial actions and their temporal segments in nearly frontal- view face image sequences",Proc. IEEE conf. Systems, Man and Cybernetics, vol. 4, pp. 3358- 3363 (2005).

[65] W. Zheng, X. Zhou, C. Zou, L. Zhao.: "Facial Expression Recogn i tion Using Kernel Canonical Correlation Analysis (KCCA)",IEEE Trans. Neural Networks, vol. 17, no. 1, pp. 233 238 (2006).

[66] K. Anderson and P.W. McOwan.: "A Real- Time Automated System for Recognition of Human Facial Expressions",IEEE Trans. Systems, Man, and Cybernetics Part B, vol. 36, no. 1, pp. 96- 105 (2006).

[67] P.S. Aleksic, A.K. Katsaggelos.: "Automatic Facial Expression Recognition Using Facial Animation Parameters and MultiStream HMMs",IEEE Trans. Information Forensics and Security, vol. 1, no. 1, pp. 3- 11 (2006).

[68] M. Pantic and I. Patras.: "Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments Form Face Profile Image Sequences",IEEE Trans. Systems, Man, and Cybernetics Part B, vol. 36, no. 2, pp. 433- 449 (2006).

[69] N. Sebe, M.S. Lew, Y. Sun, I. Cohen, T. Gevers, T.S. Huang.: "Authentic Facial Expression Analysis",Image and Vision Computing, vol. 25, pp. 1856-1863 (2007).

[70] I. Kotsia and I. Pitas.: "Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines",IEEE Trans. Image Processing, vol. 16, no. 1, pp. 172- 187 (2007).

[71] J. Wang and L. Yin.: "Static Topographic Modeling for Facial Expression Recognition and Analysis",Computer Vision and Image Understanding, vol. 108, pp. 19 34 (2007).

[72] F. Dornaika and F. Davoine.: "Simultaneous Facial Action Tracking and Expression Recognition in the Presence of Head Motion",Int. J. Computer Vision,vol. 76, no. 3, pp. 257 281 (2008).

[73] I. Kotsi, I. Buciu, I. Pitas.: "An Analysis of Facial Expression Recognition under Partial Facial Image Occlusion",Image and Vision Computing, vol. 26, no. 7, pp. 1052- 1067 (2008).

[74] R.G.Shoja, O.Arandjelović, A.Bendada, X.Maldague.: "Infrared face recognition: a comprehensive review of methodologies and databases", Pattern Recognition, vol. 47, no. 9, pp. 2807-2824 (2014).

[75] Wikipedia, `http://en.wikipedia.org/wiki/Infrared_vision`

[76] X. Maldague.: "Theory and Practice of Infrared Technology for Non Destructive Testing", John-Wiley and Sons (2001).

[77] S.G.Kong, J.Heo, B.R. Abidi, J.Paik, M.A. Abidi.: "Recent advances in visual and infrared face recognitiona review", Computer Vision and Image Understanding, Vo. 97, no. 1, pp. 103-135 (2005).

[78] R.S.Ghiass, O.Arandjelović, A..Bendada, X.Maldague.: "Recent advances in visual and infrared face recognitiona review", Computer Vision and Image Understanding, Vo. 97, no. 1, pp. 103-135 (2005).

[79] F.H.Netter and J.T.Hansen.: "Atlas of Human Anatomy", 3rd edition, California: ICON Learning Systems (2002).

[80] C.Starr and B.McMillan.: "Human Biology, California", Books Cole Thomson Learning (2003).

[81] T.Ogasawara, Y.Kitagawa, T.Ogawa, T.Yamada,Y.Kawamura, K.Sano.: "MR Imaging and Thermography of facial angioedema: A case report", Oral Surgery Oral Medicine Oral Pathology, vol. 92, no. 4, pp. 473-476 (2001).

[82] J.G.McGimpsey,A.Vaidya,P.A.Biagioni, P.J.Lamey.: “Role of thermography is the assessment of infraorbital nerve injury after malar fractures”, British journal of Oral and Maxillofacial Surgery, no. 38, pp. 581-584 (2000).

[83] J.Pessa, V.Zadoo, P.Gerza, E.J.Adrian,A.Dewitt, J.Garza.: “Double or bifid zygomatcius major muscle: anatomy,incidence and clinical correlation”, Journal of Clinical Anatomy, vol. 11, pp. 310-313 (1998).

[84] B.F.Jones.: “A reappraisal of use of infrared thermal image analysis in medicine”, IEEE Transactions on Medical Imaging, vol. 17, no. 6, pp. 1019-1027 (1998).

[85] J.T.Cacioppo, L.K.Bush, L.G.Tassinary.: “Microexpressive facial actions as function of affective stimuli: replication and extension”, Personality and Social Psychology Bulletin, vol. 18, pp. 515-526 (1990).

[86] P.Buddharaju, I.T.Pavlidis, P.Tsiamyrtzis, M.Bazakos.: “Physiology-based face recognition in the thermal infrared spectrum”, Pattern Analysis and Machine Intelligence, IEEE Transactions, vol.29, no. 4, pp. 613-626 (2007).

[87] F.J.Prokoski, I.T.Pavlidis, P.Tsiamyrtzis, M.Bazakos.: “Method and apparatus for aligning and comparing images of the face and body from different images”, U.S. Patent No. 6,496,594 (2002).

[88] B. O.Kane, P. Sandick, T. Shaw and M. Cook.: “Dynamics of human thermal signatures”, Proceedings of the Inframation Conference (2004).

[89] S.Jarlier, D.Grandjean, S. Delplanque, K.N.Diaye, I. Cayeux, M.I.Velazco, D.Sander, P.Vuilleumier,K.R. Scherer.: “Thermal analysis of facial muscles contractions”, Affective Computing, IEEE Transactions , vol. 2, pp. 2-9, (2011).

[90] Y.Yoshitomi.:“ Facial expression recognition for speaker using thermal image processing and speech recognition system”, Proceedings of the 10th WSEAS International Conference on Applied Computer Science, pp. 182-186 (2010).

[91] Koda, Y., Yoshitomi, Y., Nakano, M., Tabuse, M.:“ A facial expression recognition for a speaker of a phoneme of vowel using thermal image processing and a speech recognition system”, The 18th IEEE International Symposium on Robot and Human Interactive Communication, ROMAN 2009, pp. 955-960 (2009).

[92] I. Pavlidis, N. L. Eberhardt, J. A. Levine.:“ Human behaviour: Seeing through the face of deception”, Nature, vol. 415, pp. 35-35, 01/03 (2002).

[93] J. A. Levine, I. Pavlidis, M. Cooper.:“ The face of fear”, Lancet, vol. 357,pp. 1757 (2001).

[94] N.Powar.:“ An approach for the extraction of thermal facial signatures for evaluating threat and challenge emotional states”, Doctor of Philosophy in Electrical Engineering, The School of Engineering of the university of Dayton (2013).

[95] Y. Tian, T. Kanade, J. Cohn .: “Facial Expression Analysis”, Handbook of Face Recognition, Springer New York, pp. 247-275 (2005).

[96] Z.Zeng, M.Pantic, G.T.Roisman, T.S.Huang.: “A survey of affect recognition methods: Audio, visual, and spontaneous expressions”, IEEE Trans. Pattern Anal. Mach. Intell, vol. 31, no. 1, pp. 39-58, (2009).

[97] B.Fasel and J.Luettin.: “Automatic facial expression analysis: a survey”, Pattern Recognition, vol. 36, pp. 259-275 (2003).

[98] R.Nakanishi, K.I.Matsumura.: “Facial skin temperature decreases in infants with joyful expression”, Infant Behavior and Development, vol. 31, p. 137-144 (2008)

[99] T. Kanade, J. Cohn, Y. Tian.: “Comprehensive Database for Facial Expression Analysis”, Proc. IEEE Intl Conf. Face and Gesture Recognition, pp. 46- 53 (2000)

[100] http://vasc.ri.cmu.edu//idb/html/face/facial_expression/index.html

[101] http://www.mmifacedb.com/

[102] http://emotion-research.net/toolbox/toolboxdatabase.2006-09-28.5469431043

[103] http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html

[104] A.M. Martinez and R. Benavente.: “The AR Face Database”, CVC Technical Report (1998)

[105] http://www.ri.cmu.edu/research_project_detail.html?project_id=418&menu_id=261

[106] http://www.kasrl.org/jaffe.html

[107] M.J. Lyons, S. Akamatsu, M. Kamachi, J . Gyoba.: “Coding Facial Expressions with Gabor Wavelets”,Proc. 3rd IEEE Int . Conf. on Automatic Face and Gesture Recognition, pp. 200- 205 (1998).

[108] M.Pantic, M.S.Bartlett.: “Machine Analysis of Facial Expressions”,Face Recognition, I-Tech Education and Publishing, Vienna, Austria, pp. 377-416 (2007).

[109] NIST Equinox database, www.equinoxsensors.com/products/HID.html

[110] IRIS database, http://www.cse.ohio-state.edu/otcbvs-bench/Data/02/download.html

[111] S.Wang,Z.Liu, S.Li, , Y.Lv, , G.Wu,P.Peng, F.Chen, X.Wang.: “A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference”, IEEE Transactions on Multimedia, vol. 12, no. 7, pp. 682-691 (2010).

[112] H.Nguyen, K.Kotani, F.Chen, B.Le.: “A thermal facial emotion database and its analysis”, In Image and Video Technology,Springer Berlin Heidelberg, pp. 397-408 (2014).

[113] D.T.LIN.:“ Facial Expression Classification Using PCA and Hierarchical Radial Basis Function Network”, Journal of Information Science and Engineering, vol. 22, pp. 1033-1046 (2006).

[114] T.Kurozumi, Y.Shinza, Y.Kenmochi, K.Kotani.: "Facial Individuality and Expression Analysis by Eigenspace Method Based on Class Features or Multiple Discriminant Analysis", ICIP (1999).

[115] T.Yabui, Y.Kenmochi, K.Kotani.:" Facial expression analysis from 3D range images; comparison with the analysis from 2D images and their integration", ICIP (2003).

[116] M.Antonini, M.Barlaud, P.Mathieu, I.Daubechies, I.:" Image coding using wavelet transform", IEEE Trans. Image Processing, vol. 1, pp.205 -220 (1992).

[117] H.Nguyen, F.Chen, K.Kotani, and B.Le (2014), *Human emotion estimation using wavelet transform and t-ROIs for fusion of visible images and thermal image sequences*, Computational Science and Its Applications ICCSA 2014, Lecture Notes in Computer Science, Vol.8584, pp. 224-235.

[118] M. Turk and A. Pentland (1991), *Eigenfaces for recognition*, Journal of Cognitive Neuroscience, Vol.3, pp.71-86.

[119] M.Pantic, S.Member and L.J.M.Rothkrantz (2000), *Automatic analysis of facial expressions: The state of the art*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, pp. 14241445.

[120] J.J. Gross and R.W. Levenson, *Emotion Elicitation Using Films,* Journal of Cognition and Emotion, vol.9, no.1, pp.87-108.

[121] Image feature extraction by Gabor wavelet transforms, `www.brain.kyutech.ac.jp/~morie/topics/gabor_en.shtml`

[122] A. Koutlas and D. I. Fotiadis (2008), *An automatic region based methodology for facial expression recognition,* in IEEE Int'l Conf. on Syst. Man and Cybernetics, pp.662-666.

[123] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham (1995), *Active Shape Models-Their Training and Application,* Computer Vision and Image Understanding, Vol. 61, pp. 38-59.

[124] J. Daugman (2003) *Demodulation by Complex-Valued Wavelets for Stochastic Pattern Recognition,* Int',l J. Wavelets, Multiresolution and Information Processing, vol. 1, no. 1, pp. 1-17.

[125] T. Ojala, M. Pietikainen and T. Maenpaa (2002), *Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns,* IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987.

[126] C. Cortes and V. Vapnik (1995), *Support-vector network*, Machine Learning, vol. 20, pp.273 -297.

[127] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman (1997), *Eigenfaces vs. fisherfaces: recognition using class specific linear projection,* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.19, pp. 711720.