Investigation of Nonlinear Feature Extraction Techniques for Facial Emotion Recognition

by

Hasimah Ali
(1140610629)

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

School of Mechatronic Engineering
UNIVERSITI MALAYSIA PERLIS

2016

UNIVERSITI MALAYSIA PERLIS

DECLARATION OF THESIS

Author's full name : HASIMAH ALI

Date of birth : 03 DECEMBER 1980

Title : INVESTIGATION OF NONLINEAR FEATURE EXTRACTION TECHNIQUES FOR FACIAL EMOTION RECOGNITION

Academic Session : 2015/2016

I hereby declare that the thesis becomes the property of Universiti Malaysia Perlis (UniMAP) and to be placed at the library of UniMAP. This thesis is classified as :

☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*

☐ **RESTRICTED** (Contains restricted information as specified by the organization where research was done)*

☐ **OPEN ACCESS** I agree that my thesis is to be made immediately available as hard copy or on-line open access (full text)

I, the author, give permission to the UniMAP to reproduce this thesis in whole or in part for the purpose of research or academic exchange only (except during a period of _____ years, if so requested above).

Certified by:

SIGNATURE | **SIGNATURE OF SUPERVISOR**

(NEW IC NO. / PASSPORT NO.) | **NAME OF SUPERVISOR**

Date :_________________ | Date : _________________

NOTES : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentially or restriction.

ACKNOWLEDGMENT

First and foremost, I thank God the Almighty, Allah Subhana Wata'ala, for giving His blessing to me to complete this fruitful journey of doctoral study. This research is never solitary and thus, I have many people to thank for making the writing of this thesis bearable and to reach the end of my battle successfully.

I should thank the Ministry of Higher Education Malaysia and my employer, Universiti Malaysia Perlis (UniMAP) for giving this golden opportunity and financial support, without which I may have not pursued this study diligently. Also, I would like to acknowledge the Vice Chancellor of University Malaysia Perlis (UniMAP), Brig. Jeneral Dato' Prof. Dr. Kamaruddin Hussin, the Dean and staffs of Mechatronic Engineering of UniMAP and the Dean and staffs of Centre for Graduate Studies of UniMAP for their support and encouragement throughout my candidature.

Especially, I would like to thank my main supervisor, Dr. Hariharan Muthusamy for supervising my doctorate study. He has been very patient and dedicated with the writing and eventual completion of this thesis. His inputs and comments were tremendous and constructive and his insight proved to be very valuable.

My deepest gratitude also goes out to Prof. Dr. Sazali Yaacob, my co-supervisor, especially during the first year of my candidature, who graciously set aside time to entertain my fearful emotion of the dark ages and guide me through the literature review of this study. I deeply appreciate his genuine concern for my well-being, his true belief in me, his steady support and his kind words of encouragement.

I wish to also thank my second co-supervisor, Prof. Dr. Hamid Adom for his advice and support. Also, many thanks go to fellow members of the Intelligent Signal Processing Cluster of UniMAP, especially to my kind friends, Hazadura, Nadia, who, closely share their knowledge and guide me very well through real examples.

It is no denying that I went through a very stressful time during my study. During these times of enormous difficulty, my family was the most that I sacrificed. My sincerest and utmost gratitude go to my beloved husband, Mohamed for his kindness, allegiance and unconditional love and support by taking finest care of our children, Muriyah and Fatimah and for helping me with the household chores. I am also grateful to my beloved mother, for her patience, love, understanding and endless prayers towards my success.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	Sum of Amplitude
ANOVA	Analysis of Variance
CK	Cohn-Kanade
D	Diagonal
DWT	Discrete Wavelet Transform
ELM	Extreme Learning Machines
EMD	Empirical Mode Decomposition
ERM	Empirical Risk Minimization
FACS	Facial Action Coding System
FLD	Fisher’s Linear Discriminant
H	Horizontal
HCI	Human-Computer Interaction
HLAC	High Order Local Autocorrelation Coefficient
HOS	Higher Order Spectra
IMFs	Intrinsic Mode Functions
JAFFE	Japanese Female Facial Expression
KLFDA	Kernel Local Fisher Discriminant Analysis
k-NN	k-Nearest Neighbour
LBP	Local Binary Patterns
LDA	Linear Discriminant Analysis
LFDA	Local Fisher Discriminant Analysis
LGBP	Local Gabor Binary Patterns
LPP	Locality-Preserving Projection

MANFIS	Multiple Adaptive Neuro-Fuzzy Inference System
M_{ave}	Mean Magnitude
MFNs	Multilayer Feedforward Networks
NBCE	Normalized Bispectral Cubic Entropy
NBE	Normalized Bispectral Entropy
NBSE	Normalized Bispectral Squared Entropy
PCA	Principal Component Analysis
RBF	Radial Basis Function
SLFNs	Single Hidden Layer Networks
SRM	Structural Risk Minimization
SVM	Support Vector Machines
V	Vertical

LIST OF SYMBOLS

$f(x, y)$	2D image
$\psi(k, x)$	2D Gabor function
σ	Sigma
c_1^x	1st order cumulant
$c_2^x(\tau_1)$	2nd order cumulant
$c_3^x(\tau_1, \tau_2)$	3rd order cumulant
c_n^x	*n*th order cumulant
m_1^x	1st order moment
$m_2^x(\tau_1)$	2nd order moment
$m_3^x(\tau_1, \tau_2)$	3rd order moment
m_n^x	*n*th order moment
γ_2	Variance
γ_3	Skewness
γ_3	Kurtosis
$B(f_1, f_2)$	Bispectrum
$E[.]$	Expectation operator
$g(s, \theta)$	Radon transform
s	Small distance to the origin
θ	Angle
$E_u(t)$	Upper envelope

$E_l(t)$	Lower envelope
μ	Mean
S	Covariance
λ	Eigenvalue
v	Eigenvector
S_W	Within-class scatter
S_B	Between-class scatter
$\tilde{S}_W$	Local within-class scatter
$\tilde{S}_B$	Local between-class scatter
$A_{i,j}$	Similarity matrix
$\tilde{A}_{i,j}^W$	Weight matrices of the local within-class adjacency graph
$\tilde{A}_{i,j}^B$	Weight matrices of the local between-class adjacency graph
$K_{i,j}$	Kernel matrix
$\langle \cdot,\cdot \rangle$	Inner product
Δ	Non-redundant region
ξ	Slack variable
α_i	Lagragian multiplier
H	Hidden layer output matrix
$H^\dagger$	Moore-Penrose generalized inverse of H

Kajian Terhadap Teknik Pengekstrakkan Tak Linear bagi Pengenalan Emosi Wajah

ABSTRAK

Sejak beberapa dekad yang lalu, pengenalan emosi wajah telah menerima minat yang ketara di kalangan penyelidik dalam bidang penglihatan komputer, pengenalan corak dan bidang yang berkaitan. Peningkatan aplikasi terhadap pengenalan emosi muka telah menunjukkan kesan yang besar dalam beberapa bidang termasuklah daripada psikologi kepada interaksi manusia-komputer (HCI). Walaupun pengenalan emosi wajah telah mencapai tahap kejayaan tertentu, namun prestasinya masih jauh dari persepsi manusia. Banyak pendekatan telah dicadangkan dalam kesusasteraan. Malah, keupayaan pengenal emosi muka untuk beroperasi secara automatik sepenuhnya dengan ketepatan yang tinggi masih mencabar disebabkan masalah-masalah seperti perbezaan antara kelas, persamaan antara kelas dan perubahan halus ciri-ciri wajah. Masalah ini turut dikhuatiri apabila *physiognomies* muka berkaitan dengan umur, etnik dan jantina berbeza dari individu yang lain, seterusnya meningkatkan kesukaran mengiktiraf emosi muka. Untuk menyelesaikan masalah ini, tesis ini bertujuan untuk membangunkan teknik pengekstrakan tak linear dengan menggunakan Perintah spektral Tinggi (HOS) dan Empirikal Mod Penguraian (EMD) secara berasingan dalam mengenal tujuh emosi muka (marah, jijik, takut, gembira, neutral, sedih dan terkejut) berdasarkan imej- imej statik. Langkah awal pra-pemprosesan adalah untuk mengasingkan kawasan muka daripada imej wajah asal dengan menggunakan pengecaman muka. Imej wajah 2-D kemudian diunjurkan ke dalam 1-D isyarat muka dengan unjuran berturut-turut melalui pengubah Radon. Pengubah Radon adalah translasi dan putaran tak berubah, oleh itu ia mengekalkan variasi dalam keamatan piksel. Isyarat muka yang menggambarkan emosi diekstrak menggunakan HOS dan EMD untuk mendapatkan satu set ciri-ciri yang ketara. Dalam rangka kerja HOS, statistik tertib ketiga atau *bispectrum* yang menangkap kontur (bentuk) dan maklumat tekstur telah digunakan pada isyarat muka. Dalam kajian ini, satu set ciri-ciri bispectral digunakan untuk menghuraikan ciri-ciri ketujuh-tujuh kelas emosi. Sementara itu, dalam rangka kerja EMD, isyarat muka telah diurai menggunakan EMD untuk menghasilkan satu set kecil fungsi mod intrinsik (IMFs) melalui proses saringan. Ciri-ciri IMF yang mempamerkan corak yang unik telah digunakan untuk membezakan emosi-emosi wajah. Dalam usaha untuk mengurangkan dimensi tinggi ciri-ciri HOS dan EMD, tiga teknik pengurangan dimensi telah digunakan: Analisis Pembezaan Linear (LDA), Analisis Pembezaan Fisher Tempatan (LFDA) dan *Kernel* LFDA (KLFDA). Ciri-ciri yang diperolehi kemudian dimasukkan kepada pengelas mesin pembelajaran yang berbeza seperti k-jiran terdekat (k-NN), Mesin Sokongan Vektor (SVM) dan Extreme Mesin Pembelajaran (ELM) untuk mengklasifikasikan tujuh emosi wajah. Untuk menilai keberkesanan kaedah yang dicadangkan, dua pangkalan data penanda aras telah digunakan iaitu Ekspresi Muka Perempuan Jepun (Jaffe) dan Ekspresi Muka Cohn-Kanade. Keputusan eksperimen menunjukkan bahawa kaedah yang dicadangkan bukan sahaja menunjukkan keputusan yang lebih hebat berbanding dengan beberapa algoritma yang sedia ada tetapi juga dapat menangani imej separa tertutup serta imej bising.

INVESTIGATION OF NONLINEAR FEATURE EXTRACTION TECHNIQUES FOR FACIAL EMOTION RECOGNITION

ABSTRACT

Over the last decades, facial emotion recognition has received a significant interest among researchers in areas of computer vision, pattern recognition and its related field. The increasing applications of facial emotion recognition have shown a sizeable impact in many areas ranging from psychology to human-computer interaction (HCI). Although facial emotion recognition has achieved a certain level of success, however its performance is far from human perception. Many approaches have been constantly proposed in the literature. In fact, the ability of facial emotion recognition to operate in fully automated with high accuracy remains challenging due to various problems such as intra-class variations, inter-class similarities and subtle changes of facial features. The adhered problem is further hampered as physiognomies of faces with respect to age, ethnicity and gender, thus increase the difficulties of recognizing the facial emotion. In order to resolve this problem, this thesis aims to develop nonlinear features extraction techniques of using Higher Order Spectra (HOS) and Empirical Mode Decomposition (EMD) separately in recognizing the seven facial emotions (anger, disgust, fear, happiness, neutral, sadness and surprise) from static images. A pre-processing step of isolating face region from different background was first employed by means of face detection. The 2-D facial image was then projected into 1-D facial signal by successive projection via Radon transform. Radon transform is translation and rotation invariant, hence preserves the variations in pixel intensities. The facial signal that describes the expression was extracted using HOS and EMD to obtain a set of significant features. In HOS framework, the third order statistic or *bispectrum* that captures contour (shape) and texture information was applied on facial signal. In this work, a new set of bispectral features was used to characterize the distinctive features of seven classes of emotion. While, in EMD framework, the facial signal was decomposed using EMD to produce a small set of intrinsic mode functions (IMFs) via sifting process. The IMF features which exhibit the unique pattern were used to differentiate the facial emotions. In order to reduce high dimensionality of HOS- and EMD features, three dimensionality reduction techniques were adopted: Linear Discriminant Analysis (LDA), Local Fisher Discriminant Analysis (LFDA) and Kernel LFDA (KLFDA). The obtained features were then fed to different machine learning classifiers such as k-nearest neighbor (k-NN), Support Vector Machines (SVM) and Extreme Learning Machines (ELM-RBF) for classifying the seven facial emotions. To evaluate the effectiveness of the proposed method, two benchmark databases are used namely, Japanese Female Facial Expression (JAFFE) and Cohn-Kanade Facial Expression Database (CK). Experimental results show that the recognition rate of HOS + KLFDA + ELM-RBF and IMF1 + KLFDA + ELM-RBF have achieved the accuracy of 99.26% and 99.75%, respectively. Therefore, the proposed method not only demonstrates the superior results compared with some existing algorithms but also satisfactorily deal with partially occluded images as well as noisy images.

CHAPTER 1

INTRODUCTION

1.1 Research Background

Facial emotion or synonymously facial expression refers to the explicit transformation of human face due to the automatic response of the emotional states. Facial expressions in fact are induced by the activations of facial muscles, which result in temporally deformed permanent facial features (eyebrows, eyelids, mouth and nostril) and transient facial features (furrow and bulges). Temporal dynamics of muscular activities are typically brief, ranging from 5 seconds or less 250 milliseconds (Fasel & Luettin, 2003). A study conducted by psychologist (Mehrabian, 1968) on communications pertaining to feelings and attitudes shows that 7% of meaning resides in the spoken word, 38% in vocal utterance (the way that the word are said) and the other 55% in facial expression. This implies that facial expression forms a major part of human communication. Hence, facial expression plays an important role in human face-to-face interactions in delivering messages or intentions.

Cross-cultural research in facial expression (Ekman, 1972, 1992; Ekman & Oster, 1979; Keltner & Ekman, 2000) has shown that the six facial expression of emotions such as *anger*, *disgust*, *fear*, *happiness*, *sadness* and *surprise* are universal across human ethnicities and cultures. Ekman & Friesen (1978) have introduced Facial Action Coding System (FACS) as a tool for measuring and describing facial muscular activity.

The FACS refers to the description of an individual or combine facial muscles and tongue generated from analysis of facial anatomy. They measure the facial appearance changes in terms of FACS with 44 different action units (muscle actions) that produce them. Therefore, FACS provides a tool for behavioral science research, cognitive process and also becomes a strong basis for the development towards human-computer interaction systems.

Recent advances in image analysis, pattern recognition and computer vision open up a window towards automatic detection and classification of facial emotion. The automatic facial emotion analysis could bring facial emotion into human-machine interactions as a new modality and makes the interaction tighter and more efficient. Therefore, analysis of facial expression would be highly beneficial for fields as diverse as in behavioral sciences, medicines, physiology, security and computer science (Pantic & Rothkrantz, 2000a).

Nowadays, with advanced technology, facial emotion recognition (FER) is directed into the development of robust human-computer interface (HCI). This HCI has focused on the invention of social welfare robot that can assist the physically disabled who are bedridden or wheelchair bound to gain mobality. As the robot becomes part of our living space, it is important for a robot to understand human's mood and emotion so that a better understanding and interaction between humans and machine can be achieved. For instance, the invention of nurse robot could tend to a patient in hospital via estimating his/her emotion and create an appropriate response to the particular emotions displayed.

Dai et al. (2001) proposed a new method for monitoring patients in bed by utilizing the analysis of FER. (Breazeal, 2003) developed an expressive humanoid robot called *Kismet* that is able to perceive natural social cues either from visual or auditory channel and delivers the social signals to the man through facial expression. (Anderson & McOwan, 2004) developed *EmotiChat* application that capture the expression (e.g. *happy*) and automatically insertion emotion icon "*happy*" into the text in the *EmotiChat* application without typing.

In behavioral sciences and medicine for instance, facial expression has been used for pain monitoring system in the work of (Hammal & Kunz, 2012). Their system utilizes facial expression to help patients who are unable to convey pain in words (e.g. newborn baby or persons with serious cognitive impairments like autism). Recently, facial expression was used to generate the differences between genuine and simulated pain. (Littlewort, Bartlett, & Lee, 2009) introduced a system that was able to differentiate between real and faking with better accuracy by tracking patterns of the subtle muscle movement from subject faces.

Due to the sizable impact of FER in daily life, various approaches have been proposed in the literature. Although FER has reached a certain level of success (Deng, Jin, Zhen, & Huang, 2005; Donato, Bartlett, Hager, Ekman, & Sejnowski, 1999a; Feng, Pietikäinen, & Hadid, 2005; Gu, Xiang, Venkatesh, Huang, & Lin, 2012; Owusu, Zhan, & Mao, 2014; Pantic & Rothkrantz, 2000b; Shan, Gong, & McOwan, 2009; Shih, Chuang, & Wang, 2008; Zhao & Zhang, 2011) development of a robust FER is still ongoing and challenging as there are still many unsolved aspects due to various unpredictable facial variations and complicated exterior environment conditions. The problems such as intra-class variations which exist in facial expression images of the same type, inter-class similarities in facial expression images and subtle change of nonlinear facial features make it difficult to pre-locate facial regions and perform robust and accurate feature extraction.

The problem is further compounded by physiognomies of faces that vary from individual to another, thus making the recognition harder. Many efforts (Pantic & Rothkrantz, 2000a) have been made to deal with these variations in FER. Ideally, facial features should be robust to intra-class variations such as small amounts of translation, rotation, spatial scale and additive noise. Higher order spectra (HOS) features and empirical mode decomposition (EMD) features are suitable from this perspective. Therefore, the main objective of this research is to propose nonlinear feature extraction techniques (using HOS and EMD) to improve robustness and performances in FER system.

HOS offer some advantages in identifying non-linear coupling, Gaussianity deviation and features obtained from it can be invariants to rotation, translation and scaling (Chandran & Elgar, 1993). These features can be used for various practical applications. Thus, it motivates this research to further investigate HOS techniques on facial emotion recognition.

EMD is a multi-resolution technique suitable for nonlinear data. It decomposes complicated signals into frequency components so-called intrinsic mode functions (IMFs) (Huang et al., 1998). The EMD offers the benefits in which the basis functions can be directly derived from the signal itself based on the local characteristic time scale of the signal which provides full data-driven approach (Nunes et al., 2003) and often brings not only high decomposition efficiency but also sharp frequency and time localizations (Qing et al., 2010).