AUTOMATED MARKER PLACEMENT BASED REAL-TIME FACIAL EMOTIONAL EXPRESSION RECOGNITION SYSTEM

by

VASANTHAN A/L MARUTHAPILLAI

(1140610606)

A thesis submitted in fulfillment of the requirement for the degree of

Doctor of Philosophy

School Of Mechatronics Engineering

UNIVERSITI MALAYSIA PERLIS

2017

ACKNOWLEDGEMENT

Throughout the fulfilment of the project, I faced a lot of challenges which gave me new experiences. I can no longer expect to be spoon-fed and must take my initiative towards improvement and learning. Fortunately, some kind people are continuously supporting me in completing this project. I would not have been able to complete my PhD research if not due to the help and guidance are given to me by various people in UniMAP and also outside UniMAP.

Firstly, I would like to express my deepest thank you to God, who gave me strength and knowledge to perform the project well. I would like to express my heartfelt gratitude to my supervisor, Dr M. Murugappan and my co-supervisor, Prof. Madya Dr Wan Khairunizam Bin Wan Ahmad and Dr Shafriza Nisha bin Basah for their superb guidance that spurred me on to learn more and more to gain as much knowledge as I can. They had taught me a lot of new things which are beyond my studies presently, and I appreciate their never-ending support towards me. They never say no when I approached them, and it gave me self-belief and enthusiasm to complete the project successfully.

I would like to say my special thank you wishes to my family, relatives and friends for their encouragement and wishes which geared me up to face any obstacle in this project. I would like to express my sincere thank you to Prof Dr.Nagarajan, Dr.Gopinath Subash and my brothers, Mr. Allan Melvin, Mr. Pubalan and Mr. Murugesh for their tolerance and never-ending support from day one of the project.

Last but not least, my sincere thanks to everyone who contributed directly or indirectly in fulfilling this project and the report for their help and support. Thank you!

TABLE OF CONTENTS

CHAPTER 4 EXPERIMENTAL RESULTS AND DISCUSSIONS

CHAPTER 5 CONCLUSION

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AdaBoost	Adaptive Boosting
ADC	Analogue-To-Digital Converter
AgT	The Angle of The Triangle
AIC	Area of the Inscribed circle Triangle
AOC	Area of the Circumscribed circle Triangle
AT	Area of a Triangle
AU	Action Units
BiN	Binary Normalized Data
BG1	Background 1 (black background)
BG2	Background 2 (coloured poster background)
BpN	Bipolar Normalized Data
BU-3DFE	Binghamton University 3D Facial Expression
BS2p	Basic Stamp 2p
CCD	Charge Coupled Device
CCTV	Closed-Circuit Television
CCW	Counter Clockwise
CK+	Cohn-Kanade AU-Coded Facial Expression Database
cm	centimetre
CMD	Changes in Marker Distance
CMOS	Complementary Metal Oxide Semiconductor
CW	Clockwise
FACS	Facial Action Coding System
HMI	Human Interface Machine

HMM	Hidden Markov Models
I/O	Input/ Output
IR	Infrared
ISFER	Integrated System For Facial Expression Recognition
JAFFEE	Japanese Female Facial Expression
KNN	K-Nearest Neighbour Classifier
LCD	Liquid Crystal Display
lx	Luminous Intensity
m	meter
mAh	milli Ampere hour
MD	Marker Distance
MoBIC	Mobility of Blind and elderly people Interacting with Computers
FM	Frequency Modulation
OpenCV	Open Source Computer Vision
PCA	Principal Component Analysis
PDB	Professional Development Board
PIC	The Perimeter of the Inscribed circle Triangle
PNN	Probabilistic Neural Network
POC	The Perimeter of the Circumscribed circle Triangle
PT	Perimeter of a Triangle
PWM	Pulse Width Modulation
RBF	Radial Basis Function
RM	Ringgit Malaysia (Malaysian Ringgit)
RMS	Root Mean Square
ROC	Receiver Operating Characteristic

RoVI	Robot for Visually Impaired
SI	System International
SVM	Support Vector Machine Classifier
USB	Universal Serial Bus
VCR	Video Cassette Recorder
VGA	Video Graphics Array
WN	Without Normalized the Data
3D	3-Dimention

LIST OF SYMBOLS

$\circ$	Degree
(x, y)	Coordinate
θ	Angle
γ	Kernel Parameter
π	Pi
C	Penalty Parameter
σ	Spread Value
K	KNN parameter
N	Total number of the data
σ^2	Variance
X_{rms}	Root Mean Square
$\bar{X}$	Mean
p_e1	Right eye marker at the angle of 45^o
p_e2	Left eye marker at the angle of 45^o
p_e3	Right eye marker at the angle of 65^o
p_e4	Left eye marker at the angle of 65^o
p_m1	Right mouth marker
p_m2	Left mouth marker
p_m3	Upper mouth marker
p_m4	Lower mouth marker
p_m6	Right cheek marker
p_m7	Left cheek marker
e1	Distance between centre marker to 'p_e1' marker
e2	Distance between centre marker to 'p_e2' marker
e3	Distance between centre marker to 'p_e3' marker

e4	Distance between centre marker to 'p_e4' marker
e12	Distance between 'p_e1' marker to 'p_e2' marker
e34	Distance between 'p_e3' marker to 'p_e4' marker
m1	Distance between centre marker to 'p_m1' marker
m2	Distance between centre marker to 'p_m2' marker
m3	Distance between centre marker to 'p_m3' marker
m4	Distance between centre marker to 'p_m4' marker
m5	Distance between 'p_m1' marker to 'p_m2' marker
m6	Distance between centre marker to 'p_m6' marker
m7	Distance between centre marker to 'p_m7' marker
m13	Distance between 'p_m1' marker to 'p_m3' marker
m14	Distance between 'p_m1' marker to 'p_m4' marker
m23	Distance between 'p_m2' marker to 'p_m3' marker
m24	Distance between 'p_m2' marker to 'p_m4' marker
m34	Distance between 'p_m3' marker to 'p_m4' marker
m67	Distance between 'p_m6' marker to 'p_m7' marker
(X_{p_e1}, Y_{p_e1})	Coordinate of right eye marker at the angle of 45^o
(X_{p_e2}, Y_{p_e2})	Coordinate of left eye marker at the angle of 45^o
(X_{p_e3}, Y_{p_e3})	Coordinate of right eye marker at the angle of 65^o
(X_{p_e4}, Y_{p_e4})	Coordinate of left eye marker at the angle of 65^o
(X_{p_m1}, Y_{p_m1})	Coordinate of right mouth marker
(X_{p_m2}, Y_{p_m2})	Coordinate of left mouth marker
(X_{p_m3}, Y_{p_m3})	Coordinate of upper mouth marker
(X_{p_m4}, Y_{p_m4})	Coordinate of lower mouth marker
(X_{p_m6}, Y_{p_m6})	Coordinate of right cheek marker
(X_{p_m7}, Y_{p_m7})	Coordinate of left cheek marker
(X_c, Y_c)	Coordinate of centre marker

Pengesanan Emosi Manusia Automatik Melalui Unit Tindakan Muka (Facial Action Unit)

ABSTRAK

Pengesanan ekpresi manusia telah menarik beberapa penyelidik sejak beberapa dekad yang lalu dan juga kebanyakan penyelidik fokus penyelidikanya pada pengesanan ekspresi wajah di "offline" dan sangat sedikit penyelidikan tertumpu pada "online" pengesanan ekspresi wajah. Dalam usaha untuk membangunkan "real-time" sistem pengesanan ekspresi wajah bijak, tesis ini mencadangkan kaedah penempatan penanda secara automatik untuk mengklasifikasikan enam ekspresi muka asas (senyuman, kesedihan, kemarahan, ketakutan, kejijikan dan kejutan). Pada mulanya, penempatan penanda manual dijalankan untuk mengesan min kedudukan (jarak antara pusat muka ke lokasi tanda) daripada setiap tanda pada muka subjek. Kedudukan ini telah digunakan untuk mengembangkan penanda algoritma secara automatik untuk mengesan emosi wajah. Dalam eksperimen ini, subjek diminta meletakkan sepuluh penanda secara manual (empat penanda pada muka bahagian atas dan enam penanda pada muka bahagian bawah) di wajah mereka pada lokasi yang dinyatakan berdasarkan "Facial Action Coding System (FACS)". Penanda secara manual diletakkan dengan mengklik kursor pada setiap kedudukan imej muka dalam urutan video. Setiap subjek menjalani tiga ujian penempatan penanda bagi setiap ekspresi wajah emosi, dan data min dikira dari pusat muka. Suatu automatik penepatan penanda telah direka daripada data yang diperolehi dari manual penepatan penanda. Algoritma yang dicadangkan meletakkan sepuluh penanda pada muka subjek pada kedudukan yang ditakrifkan, dan kedudukan kordinat setiap penanda dihantar kepada algoritma "Optical Flow" untuk meramalkan kedudukan penanda bagi frame seterusnya. Pergerakan-pergerakan penanda untuk ekspresi muka yang berbeza telah dikaji dengan menggunakan jarak. Data tersebut dikaji dengan tiga statistik algoritma iaitu min, varians, dan root mean square Dalam tesis ini, sebanyak tujuh "features" untuk menganalisis prestasi sistem pengesanan ekspresi wajah. "Features" ini diekstrak dan diklasifikasikan dengan, "K-nearest neighbour (KNN), probabilistic neural network (PNN), support vector machine (SVM)". Tesis ini juga menganalisis prestasi tiga set numbor penanda yang berbeza (10, 8 dan 6) untuk mengesan enam ekspresi muka yang berbeza. Sejumlah sepulah, lapan dan enam penada muka telah dikelaskan dan pencapaian tertinggi dicapai oleh lapan penanda analisis iaitu 99.55% dengan mengunakan SVM klasifikasi.

Automated Marker Placement Based Real-Time Facial Emotional Expression Recognition System

Abstract

Facial expression recognition attracted several researchers over the past several decades and most of the researchers in the literature focus on facial expression recognition in "offline" and very few research works concentrated on real-time facial expression recognition. In order to develop an intelligent real-time facial expression recognition system, this thesis proposed an automated marker placement method for classifying six basic facial expressions (happiness, sadness, anger, fear, disgust and surprise) using real-time video sequence. Initially, manual marker placement was carried out to detect the mean position (distance between the centre of the face to the marker's location) of each marker on the subject's face. This position was used to expand the automated marker placement algorithm for facial emotion recognition. In this experiment, subjects were requested manually to place ten markers (four markers on the upper face and six markers on lower face) on their face in specified locations based on Facial Action Coding System (FACS). Trial and error approach devised the number of markers used for facial expression detection. Manual markers were placed by clicking the cursor at each position on the facial image in video sequence. The mean marker position distance was calculated from the centre of the face. Calculation of each marker position concerning the middle of the face via manual marker placement was then used to develop the automatic marker placement algorithm. The proposed algorithm places ten virtual markers on the subject's face on defined position, and the position of each marker is sent to optical flow algorithm for predicting the future marker position. These marker movements for different facial expressions have been investigated. A simple set of three statistical features (mean, variance, and root mean square) were extracted from the above parameters for facial expressions classification. In this thesis, a set of seven features were newly proposed to analyse the performance of facial expression recognition system. These extracted features were mapped into corresponding emotional facial expressions using three simple non-linear classifiers namely, K-nearest neighbour (KNN), probabilistic neural network (PNN), support vector machine (SVM). This thesis also analyses the performance three different set of markers (10, 8 and 6) to detect six different facial expressions. In overall, eight markers are an optimal number with higher accuracy and gave a maximum mean emotion classification rate of 99.55% using the support vector machine for the perimeter of inscribed circle feature.

CHAPTER 1

INTRODUCTION

1.1 Background

Emotions play a vital role in verbal and non-verbal communication and it is used to express the internal feelings of a human being. As of recent years, advances in intelligent autonomous system development is fast emerging in various fields of communication technology such as: ambient intelligence (Spanoudakis & Moraitis, 2015), pervasive computing (Allen et al., 2014), ubiquitous computing (Riekki et al., 2003), and emotion-aware ambient intelligence (AmE) (Acampora & Vitiello, 2013; Salmeron, 2012; Yu & Zhou, 2008; Zhou et al., 2007) for providing rapid emotion-aware mobile services that is adaptive, sensitive, and receptive in nature to a user's requirements, behavior, emotions and gestures. Distinctively, facial expressions are one of the many tools that could be utilized to investigate human emotions and has been discussed extensively in various previous studies (Dhall et al., 2013; Fridlund, 2014; Keltner et al., 2013; Pantic & Rothkrantz, 2000; Tian et al., 2001). In general, emotions are triggered in a person during both conscious and unconscious evaluations and are relevant to a goal or concern. Therefore, emotions are positive when a concern is advanced and negative when a concern is impeded.

Vision is the primary approach for acquiring information from the surroundings. Under that roof, image is classified as another form of vision, whereby different images can induce different emotions in a human. Although vision based approaches provide a higher rate of emotion recognition, it still has several limitations and challenging issues. Most researchers are focused on recognizing six basic universal emotions (happiness, sadness, surprise, fear, disgust and anger) and neutral (Ghandi et al., 2010; Majumder et al., 2014; Petrantonakis & Hadjileontiadis, 2010; Richoz et al., 2015; Velusamy et al., 2011; Zeng et al., 2009).

Presently, several methodologies have been proposed to develop a humanoid system for intelligent human-robot interaction (Mladineo et al., 2015; Sellami et al., 2015; Su et al., 2015). Nevertheless, the intelligence of a human-robot interaction highly relies on recognising human emotions in a faster and efficient manner. Based on this approach, facial expression and speech modality are considered for revealing emotional experiences. They also provide important communicative cues during social interactions. In general, a robotic emotion recognition system will enhance the interaction between human and robot in a natural manner.

Facial Action Coding System (FACS) is the most popular method used by many researchers to identify the behaviour of emotions (Li et al., 2013; Savran et al., 2010; Sun et al., 2008; Velusamy et al., 2011; Zhao & Wang, 2008). In year 1978, Ekman & Friesen developed the FACS method to analyse facial emotional behaviour. The FACS is a complete human-observer based system which is designed to detect subtle changes in facial features and fully controllable facial models by manipulating the single actions which are called Action Units (AUs). Based on FACS, facial behaviors are analyzed using 46 action units, whereby 30 action units are anatomically related to the contractions of specific facial muscles (18 AU's are for lower face, and 12 AU's are for upper face) and the remaining 16 action units are a combination of specific facial muscles (Tian et al., 2001). Ekman & Friesen (1978) discussed facial muscle activation with different emotions and defined the facial AU system for classification of facial expressions. Table 1.1 shows the effective changes of AUs in the facial muscles for each emotion developed by Ekman & Rosenberg (2005). However, most of the research works in this literature discusses the development of facial expression recognition system in a laboratory environment, and only limited studies were performed in real time scenario (Ryan et al., 2009; Suk & Prabhakaran, 2014).

Table 1.1: Action Unit studies by Ekman and Friesen

Emotions	Action Units (AU's)	FACS Name	Muscular Basis
Happiness	AU6	Cheek Raiser	Orbicularis oculi (pars orbitalis)
	AU12	Lip Corner Puller	Zygomaticus major
Sadness	AU14	Dimpler	Buccinator
	AU15	Lip Corner Depressor	Depressor anguli Oris (also known as triangulation)
Surprise	AU12	Lip Corner Puller	Zygomaticus major
	AU5B	Upper Lid Raiser	Levator palpebrae superioris, superior tarsal muscle
	AU26	Jaw Drop	Masseter; relaxed temporalis and internal pterygoid
Fear	AU12	Lip Corner Puller	Zygomaticus major
	AU4	Brow Lowerer	Depressor glabellae, corrugator supercilii
	AU5	Upper Lid Raiser	Levator palpebrae superioris, superior tarsal muscle
	AU7	Lid Tightener	Orbicularis oculi (pars palpebralis)
	AU20	Lip Stretcher	Risorius platysma
	AU26	Jaw Drop	Masseter; relaxed temporalis and internal pterygoid
Disgust	AU9	Nose Wrinkler	Levator labii superioris alaeque nasi
	AU15	Lip Corner Depressor	Depressor anguli oris (also known as triangularis)
	AU16	Lower Lip	Depressor labi inferioris
Anger	AU4	Brow Lowerer	Depressor glabellae, corrugator supercilii
	AU5	Upper Lid Raiser	Levator palpebrae superioris, superior tarsal muscle
	AU7	Lid Tightener	Orbicularis oculi (pars palpebralis)
	AU23	Lip Tightener	Orbicularis oris

There are several methods used in facial expression analysis such as Local Phase Quantization (LPQ), Pyramidal Histogram of Gradient (PHOG), Facial Action Coding System (FACS), Local Binary Patterns (LBP) and Optical Flow Algorithm (OFA). These methods were further discussed by Lonare & Jain (2013). Optical Flow Algorithm (OFA) has been widely used by many researchers to identify the AUs changes in a real-time environment for facial emotion detection (Gibson, 1950; Lonare & Jain, 2013). However, OFA is usually applied to calculate the relative motion between an observer and the scene in the motion of objects, surfaces, and edges in a visual appearance (Warren & Strelow, 1985). There are several methods for implementing the optical flow algorithm, such as phase correlation, block-based method, a discrete optimisation method, and differential method (Glocker et al., 2008). Phase correlation is used as image registration, and it estimates the relative translative offset using fast frequency-domain approach between two similar images (Robertson et al., 2014). Block-based methods are mainly used for maximising the normalised cross-correlation or minimising the sum of absolute differences or sum of squared differences to estimate the comparative moment between two images (Karasulu & Korukoglu, 2013). Differential methods are based on partial derivatives of the image pixels such as Lucas– Kanade method, Horn–Schunck method, Buxton– Buxton method, Black–Jepson method and General variation method (Lucas & Kanade, 1981).

Moreover, automated face and eye detection plays a significant role in autonomous facial expression detection. Whereby, a preeminent automated face and eye detection system should detect the user in a complex scenes, with cluttered backgrounds and also be able to locate the exact position of the user's face in the scene (Fasel & Luettin, 2003). Using face detection, facial features such as the eyes, nose and mouth serve as reference points to detect the faces (Essa & Pentland, 1997). Therefore, several face detection methods are reported in the literature for facial expression recognition (Bartneck, 2000; Suk & Prabhakaran, 2014). However, Action Unit (AU) based face detection has been referred by many research works in comparison to other methods due to the simple algorithm, lesser computational complexity and easier implementation on real-time systems (Ekman & Friesen, 1978; Zhang et al., 2008). The Viola & Jones

(2001) face detection method was used by Zhang (2008) to detect the user's face and eyes from the image. The full structure of the Viola and Jones method will be discussed in Chapter 3.

These days, researchers have started using virtual markers for facial expression detection (Zhang et al., 2008). Noticably, several number of markers were used to detect the facial expression in both laboratory and real-time environments, such as 62 markers (in a study by Kotsia & Pitas (2005), 22 markers in a study by Bajpai & Chadha (2010) and Michel & Kaliouby (2003), 21 markers in a study by Srivastava (2012) and 12 markers in a study by Ghandi et al. (2010). Most of the above works recognised six basic facial expressions (Ghandi et al., 2010; Kotsia & Pitas, 2005; Michel & Kaliouby, 2003). It is understood that virtual markers based facial expression detection offers several advantages over conventional methods such as: (i) quickly investigatses the movement of markers (when facial expression take place) (ii) users do not need to wear any special equipment or reflecting stickers on their face (iii) works in different environments and has lesser computational complexity (Ghandi et al., 2010). However, the performance of virtual makers are affected by lighting conditions and camera quality. It requires a minimum light intensity <30 lux (less than 30 lux) for better facial expression detection and higher camera resolution (Ghandi et al., 2010).

The research aim of this study is to propose a novel method for automated virtual marker placement in the subject's face for detecting six facial emotional expressions and comparing its emotion recognition performances with manual marker placement. In this study, ten virtual markers are placed automatically in a particular location on the subject's face, and a web camera (Logitech®) is used for capturing the facial emotional expression sequences during the experiment. The analysis of marker was carried out by reducing the number of markers from ten to eight and six consequently to reduce the computational complexity. Haar cascade database, which is a built-in function of Open Computer Vision (Open CV) was used to detect the subject's face from the video sequences captured by the web camera. The initial marker positions (x-y coordinates) are passed to the Lucas– Kanade optical flow algorithm to predict future marker positions. The distance from each marker from the centre point of the subjects' face is calculated. Some simple statistical features was computed from the extracted distances. Then, the extracted features were mapped into corresponding emotions using three non-linear classifiers namely K-Nearest Neighbour (KNN), Support Vector Machine (SVM) and Probabilistic Neural Network (PNN). This complete algorithm is then applied in Microsoft Visual Studio platform with an Open CV library using C++ programming language in a Desktop Computer with Intel i3 processor with 2 GB ROM in Windows operating system.

1.2 Problem Statement

Muscle movement is the root cause of facial changes which is describe as facial expressions. Facial expressions are a reaction from a person's inner feelings, emotional situation, intentions or social communications (Tian et al., 2001). Initially, facial expression investigation was principally a research subject for psychologists and behavioural scientists, since the determining research by Darwin (1948) and Ekman (2006). Generally, facial expression research investigates six basic emotions: happiness, surprise, fear, anger, sadness and disgust. These expressions are recognized when a unique prototype with original contents of facial expressions are delivered by each