Ministry of Higher Education and
Scientific Research
University of Diyala
Department of Computer Science

Improve the Performance of Recognition Facial Expression Using Speech and Image in Video

A Thesis
Submitted to the Department of Computer Science\ College of Sciences\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science

By

Meaad Hussein Abdalhadi

Supervised By

Assist.Prof. Dr. Jumana W. Salih

2020 A.D. **1442 A.H.**

بِسمِ ٱللَّهِ ٱلرَّحمنِ ٱلرَّحِيمِ

﴿ وَمِنْ آيَاتِهِ خَلْقُ السَّمَاوَاتِ وَالْأَرْضِ

وَاخْتِلَافُ أَلْسِنَتِكُمْ وَأَلْوَانِكُمْ إِنَّ فِي ذَلِكَ

لَآيَاتٍ لِلْعَالِمِينَ ﴾

صَدَقَ ٱللَّهُ العَظِيمِ

سورة الروم

الآية (٢٢)

Dedication

To...

My family

My dear parents

My dear husband

My children Hassan& Farah

All our distinguished teachers those who paved the way for our science and knowledge

✍

Meaad Hussein Abdalhadi

Acknowledgment

First of all, praise is to Allah Lord of all creation, for all the blessing was the assistance in carrying out this research until its end.

I would like to express my thanks to my supervisor, Dr. Jumana Waleed , for her supervision of this research and for generosity, patience, and constant guidance throughout the work. It was my great fortune to get advice and guidance from her. My thanks to the academic and administrative staff in the Department of Computer Science.

I would like to express my gratitude to my mother, sister, and brothers who were unlimited support and patience.

Finally, there are not enough words to thank my dear husband for his support, belief in me all the time, and his encouragement during my studies. Praise be to God who helped me and gave me the ability and strength to fulfill and fulfill my mission.

Meaad Hussein Abdalhadi

Abstract

Speech and face emotion recognition can be widely used in many applications, like assessing customer satisfaction with the quality of services in a call center, detecting/assessing the emotional state of children in care, and to recognize human emotion by the robot. There are many challenges in the speech and image face recognition systems, including recording a real dataset in a natural environment without using any filter recording device to enhance the quality of a signal. The other challenge is the ambiguity about the list/definition of emotions, the lack of agreement on a manageable set of uncorrelated speech-based emotion relevant features, and the difficulty of collected emotion-related datasets under natural circumstances.

In this thesis, to cope with these challenges a system of identifying human speech and facial emotions using a Support

Vector Machine (SVM) has been proposed to improve detection performance effectively with multiple emotions. Facial affection was detected by using the lower half of the face after extracting the important properties by the histograms of oriented gradients (HOG) algorithm, and the results obtained from the face showed a high accuracy that reached (91%) and this accuracy is high compared to the rest of the research and systems that used the entire face to be able to distinguish emotion and use many algorithms to discover features.

The emotion was detected through speech using Mel-frequency cepstral coefficients (MFCC)and pitch after extracting the important features, and the results obtained from sound showed high accuracy and reached (90%).

Contents

List of Figures

List of Tables

List of Abbreviatiations

Abbreviations	Meaning
2D	Two Dimension
AD	Analog-to-digital converter
ANN	Artificial neural network
BRIEF	Binary robust independent elementary features
CNN	Convolution Neural Network
DCT	Discrete Cosine Transform
FFT	Fast Fourier Transform
FIR	First-order high pass filter
GMM	Gaussian Mixture Models
HCI	Human-Computer Interaction
HMM	Hidden Markov Models
HOG	Histograms of oOriented Gradients
JPEG	Joint Photographic Experts Group
LBP	local binary pattern
SER	Speech Emotion Recognition
LSTM	Long Short Term Memory
MED	Minimum energy density
MFCC	Mel Frequency Cepstral Coefficient
ML	Machine Learning
MLER	Modified Low Energy Ratio
PC	Pulse Clarity
PC	Personal Computer
RBF	Radial Basis Kernel
RTER	Real-Time Emotion Recognition
SP	Spectral Centroid
SVM	Support Vector Machine
t	Threshold
ZCR	Zero-Crossing Rate
μ	Mean Value
AI	Artificial Inttelligence

Chapter One

General Introduction

Chapter One

General Introduction

1.1 Introduction

Commonly, the term of emotions is daily used. Although emotions have different definitions that are depending on the term of psychology, emotions can be defined as a complicated case of sensation which leads to behavior and physical. Generally, the emotion theory has two major categories; somatic and cognition. The first category is dependent on somatic features and sought for describing emotional expressions and their perceptions [1]. On the other hand, the second category is depended on an important component of emotion and the subjective appearance that could be unintentional or intentional, unconscious or conscious, and took a form of a thought or a judgment [2].

Within Human-Computer Interaction (HCI), the processes of emotions are inextricably joint with reasonable decisions; consequently, efficient interaction has acquired considerable interest. Thus, the emotional state of the user should be identified. Depend on the theory of psychology, there are six widely acceptable typical emotions: sadness, happiness, anger, neutral,fear, and surprise. The human speech tone and the motion of facial possess are one of the main roles to express emotions. Emotions are capable of considerably changing the sense of messages. The human facial is tending to be the most obvious form of emotional communications, however, in response to various social conditions, it is also easy to be controlled compared with speech and other types of expressions [3].

The emotional expression and human affective state recognition are necessary capabilities for human interaction and social integration. In recent years, the studies of emotion recognition have attracted the interest of researchers in diverse applications; such as human-computer interfaces, human-robot interaction systems [4][5]driver assistance, and alerting systems[6], etc. Table (1.1) shows several applications in the areas of emotion recognition.

Table (1.1): The emotion recognition systems applications.

Areas	**Applications**
Medicine	-Rehabilitation (help monitoring). -Companion (enhance realism). -Counseling (client's emotional state). -Health care (patients' feelings about treatment).
E-learning	-Adjust the presentation style of an online tutor. -Detect the state of the learner.
Monitoring	-Car driver (detect state the alert other cars).
Law Implementation	-Deeper discovery of depositions.
Marketing	-Emotion is vital in purchasing decisions.
Entertainment	-Recognize the mood and emotion of the user.

1.2 Related works

Recently, the efforts of researchers in HCI are focused on how to make the computer capable of understanding the emotions of a human. Human speech is a fundamental communication means for interaction. The speech emotion is more important as it doesn't change the speech linguistic content but alters its effectiveness. It directly affects in making a decision, cognition, perception, creativity, reasoning, memory, and attention.

The emotion recognition represents a difficult issue, especially, when the emotion recognition is accomplished via utilizing the signal of speech. Several important types of research have been presented in this field and the main faced challenges are; selecting a speech database, identifying various features regarding speech, and the suitable selection of the classification approach [7].

P. Shegokar and P. Sircar 2016, [8]proposed a speech-based emotion recognition scheme in which the selection of features is depended on the transformation of continuous wavelet and the coefficients of prosodic. In this presented scheme, various SVMs are utilized as a classification model. The experiment results show that the best rate of recognition is 60.1%.

S. Basu et al. 2017, [9]proposed a speech-based emotion recognition technique in which thirteen Mel Frequency Cepstral Coefficient (MFCC) and thirteen components of acceleration were used as features and Convolution Neural Network (CNN) with Long Short Term Memory (LSTM) as a classification approach. The obtained result of the accuracy was approximately 80%. This technique can provide better results when feeding it with a larger database. The same result of accuracy was obtained by **M. S. Likitha et al. 2017,** [10]where the MFCC features are used for feature extraction with SVM as a classification model.

Z. Han and J. Wang 2017, [11]proposed a technique of speech-based emotion recognition using SVM and Gaussian Kernel Nonlinear Proximal SVM. In this technique, the speech signal is firstly preprocessed, and then the features of speech prosody and quality are extracted. After that, SVM and Proximal SVM are utilized as a classification model for obtaining the final result of emotion recognition, where the average rate of recognition was 80.75% with SVM, and 86.75% with Proximal SVM. These obtained results show that the technique using Proximal SVM provides a better rate of emotion recognition, also, it is faster three times than SVM. this proposed technique needs to utilize more efficient features for resulting high results.

A. Bhavan et al. 2019, [12]proposed a speech emotion recognition technique based on the extraction of a set of spectral features (MFCCs and spectral centroids) that are preprocessed and reduced to the desired set of features. In this presented technique, a bagged ensemble comprising of SVMs with a Gaussian kernel was proposed to be utilized as a classification model. The obtained result of the accuracy was 84.11%. This technique is only concentrated on acoustic features, the utilizing of the linguistic features (semantic features) may work on improving the performance of the recognition technique.

On the other hand, the most popular emotion recognition approaches are based on human facial images that can help in HCI as well as several applications.

T. Kundu and C. Saravanan 2017, [13]proposed a facial emotion recognition technique focused on using artificial neural network (ANN) and SVM. In this technique, firstly, the regions of the facial (eye and mouth) are analyzed and fed into ANN. Secondly, the binary robust independent elementary features (BRIEF) descriptor is utilized for extracting the texture information, and the classification is done utilizing SVM. The obtained accuracy of this technique was 79.1%.

V. M. Álvarez et al. 2018, [14]presented a comparison between various landmark-based classifiers for Facial emotion recognition. In this work, several algorithms of face detection and alignment are applied, after that, a set of emotion-labeled landmarks are fed to various machine learning classifiers for comparing their results. The average rate of accuracy for the multiple layers of classifiers was 89%.

N. Lopes et al. 2018, [15]presented a facial based emotion recognition model to differentiate between the facial expressions of the elderly (older than 60 years) and the others (less than 60 years). In this model, Viola-Jones and Haar features were utilized for extracting the face, then, the Gabor filter is utilized for extracting the facial features to later be classified using a Multiclass SVM. The obtained average of accuracy was 80.5% for the elderly and 87.93% for the other individuals.

C. Cuong et al. 2018, [16] propose a new method to speed up the computational performance of smile detection algorithm using a specialized architecture of Faster Region ConvolutionalNeural Network (Faster R-CNN). The evaluation from GENKI-4K dataset shows that network gains up to 50% faster inference performance and 2 times faster in training than the original Faster R-CNN with an accuracy of 84.5%, which is acceptable for predicting and classifying smile from given images.

There are several limitations to only utilizing speech for recognizing emotion. Therefore, human facial expression can be combined with speech signals to obtain a considerable influence on emotion recognition results.

1.3 Problem Statement

Emotion recognition is one of the topics that have attracted much attention lastly due to its importance in many areas like the applications which require human-computer interaction (HCI).

Extract the features related to the emotional state of speech, image, and what the model that gives the best recognition remains one of the important research challenges to distinguish the system with the highest accuracy. Therefore, the main challenge in our thesis is to build a system that can distinguish the emotional state in Real-Time and compares the performance of the classifier in terms of accuracy rates.

A major problem in this system is the difficulty in dealing with two types of databases for each of the images and speech and the synchronization between them.

1.4 Aim of the Thesis

The main aim of this thesis is to recognize the human speech and facial emotion using the SVM classification algorithm in which the Mel-frequency cepstral coefficients (MFCC) and histograms of oriented gradients (HOG) descriptor are used for extracting features from the human speech and facial, respectively, to obtain high accuracy.

1.5 Outlines of the Thesis

In this section, the global structure of this thesis is submitted, and a brief description of each chapter is presented to give the reader an evident conception about the whole of the work.

Chapter Two: (Theoretical Background)

Chapter two covers the basic concepts of SVM, face recognition, sound recognition.

Chapter Three: (The Proposed System)

This chapter describes the proposed emotion recognition system with their designs and implementations.

Chapter Four : (Results and Discussion)

This chapter explains the results and tests that have been got from the proposed system.

Chapter Five: (Conclusion, and Suggestion for Future Works)

This chapter offers conclusions and suggestion systems for future works.